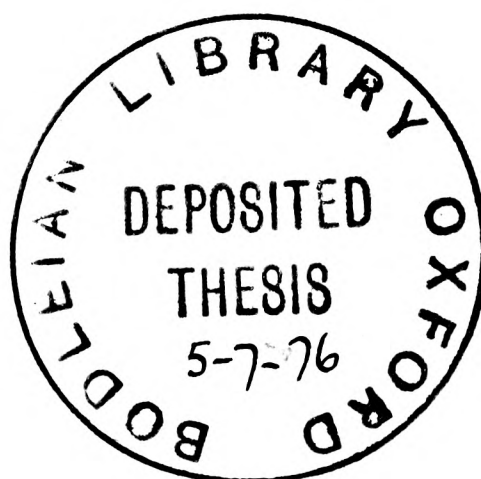


**Probability, Frequency and Evidence**

by

**Frederick C. Benenson**

**A Thesis submitted to the Board of Faculty of Literae Humaniores,  
Oxford University, for the degree of Doctor of Philosophy.**



October 1975 [e 1976] m

## ABSTRACT

Chapter I of this thesis considers the frequency theory of probability and in particular its treatment of individual events. Attempts by frequentists to give an account of ordinary statements of probability which are generally about individual events are criticized at length: Reichenbach's attempts to deal with the problem of individual events is found to be unsatisfactory in virtue of his introduction of such dubious entities as 'fictitious probabilities'; Salmon's related suggestion to treat the problem of the single case in terms of 'weights' determined by an 'application' of his theory of probability is also found to be unsatisfactory, for the concept of probability defined by his theory can not, on his own admission, be applied to the single case. Proposals by Keynes and Popper to treat the single case along frequentist lines, while apparently more promising, turn out in the end to be too rudimentary and sketchy to fulfill even the minimal condition for an adequate semantic definition of probability - the specification of a probability function which provides an interpretation of the axioms of probability. The basic difficulty for a frequentist in assigning probabilities to individual events is (not surprisingly) found in Chapter I to be that the probability of an individual event will vary depending on what reference class is chosen to determine its probability.

In Chapter II, however, the dependence of the probability of individual events on the reference class chosen is seen not to present a genuine difficulty in formulating an adequate semantic definition of probability. The variability of the probabilities assigned to individual events indicates that they are relational probabilities crucially involving a reference class in one term of the relation; once recognized the relational character of the probability of individual events provides the basis of an adequate semantic interpretation of the standard axioms of probability. It is shown that the resulting theory of probability can meet the objections raised by Salmon, Von Mises and Reichenbach against the assignment of probabilities to individual events and, indeed, this theory of probability resolves contradictions found to be inherent in Reichenbach and Salmon's treatment of the single case.

Chapter III begins with a discussion of a methodological principle by which the relational theory given in Chapter II can be applied to actual situations to determine the unique probability values often needed for the purposes of action. This is the well-known principle of choosing the narrowest reference class for which statistics are available. This principle, usually referred to in the thesis as that of choosing the narrowest available reference class, is given a preliminary (and traditional) formulation at the start of Chapter III. Von Mises' requirement of randomness is then analyzed in light of this principle and it is



found that if we employ this principle to determine unique probabilities, only random classes of events can be used to assign unique probabilities to every individual member of the class. This is found to account for the view - expressed in Von Mises' requirement of randomness - that only random classes are truly probabilistic.

In Chapter IV various problems concerning the principle of choosing the narrowest available reference class are considered. Some of these are familiar, e.g. Ayer's objection to frequentists' use of this principle, the difficulties caused by unlikeliest predicates, and the problems raised by the quite common absence of complete statistical knowledge. Less familiar problems are also considered, in particular Chapter IV is concerned with the question of precisely what kinds of reference classes can be considered 'available' for determining the probability of individual events. A solution to this problem is devised, based, in the first instance, on the concept of an effectively calculable function introduced by Church into discussions of randomness. A reference class is said to be available for determining the probability of an individual event if and only if there exists a determinate procedure which, if carried out, would yield the result that the event belongs to that reference class. In the case of reference classes formed by mathematical rules of selection, the existence of a determinate procedure leading to a result is equated with the existence of a recursive function leading to that result. In the case of reference classes based on empirical predicates, the class of determinate procedures is the totality of experimental procedures extant in the scientific field in question. This line of argument is extended to explicate the concept of 'available evidence', which figures prominently in many theories of probability. Various objections to this explication of available evidence are considered and a more refined analysis eventually emerges.

Chapter V considers Carnap's thesis that there exist two distinct concepts of probability and it is shown that this conclusion can be avoided by adopting the frequency theory for the single case given in Chapter II in place of the standard frequency theory - the definition of probability expressed in our theory is shown to be an instance of the definition of probability as a quantitative relation between evidence and hypothesis and thus to have distinct affinities to Carnap's theory of probability<sub>1</sub>. Chapter V concludes with a discussion of subjectivism in theories which make probability a relation between evidence and hypothesis and it is shown how both our theory of probability and the logical relation theory avoid such subjectivism.

Chapters VI and VII are concerned with the principle of indifference. Chapter VI begins with a survey of a variety of opinions on the principle and then presents a new interpretation of it as a rudimentary form of semantic definition of probability. It is shown that the principle, in stating conditions under which alternatives are equally probable, actually fixes identity conditions for the concept of probability and, as widely understood since Frege, any statement which fixes identity conditions for a concept fixes a particular sense for that

concept, a particular semantics. The definition of probability encapsulated in the principle is that of probability as a comparative relation of evidential support. It is further explained how such a comparative conception of probability will have led to the numerical assignments of probability traditionally arrived at by use of the principle of indifference - when, as was the case with the principle's usual employment, a set of  $n$  mutually exclusive and exhaustive hypotheses are (comparatively) judged to be equally likely, it follows directly from the axioms of probability that each hypothesis has the numerical probability  $\frac{1}{n}$ . That the principle of indifference encapsulates the semantic definition of probability as the comparative concept of evidential support is found to accord with the classical theorists' definition of probability as the ratio of favourable cases to possible cases, where each case is equi-possible. If we take the expression 'equi-possible' to be an undefined primitive term, the classical definition of probability constitutes a natural (and fruitful) uninterpreted axiomatic definition of probability. In accordance with the argument of the earlier part of Chapter VI, the principle of indifference - which even in classical times was regarded as fixing a sense or interpretation for the expression 'equi-possible' - then can be understood as providing, in a rudimentary way, a semantic interpretation for the basic undefined term of this axiom system.

Chapter VII, which concludes the thesis, analyzes a variety of objections that have been raised against the principle and the analysis there supports the over-all claim that the principle of indifference was the first, and therefore, not surprisingly, somewhat confused, attempt to state the highly plausible definition of probability as the comparative concept of evidential support.

Approximate length of thesis: 70,000 - 73,000 words.

## ACKNOWLEDGMENTS

Besides the intellectual debt I owe to various prominent philosophers of previous generations - of whom Keynes, Carnap and Reichenbach have, perhaps, influenced me the most - I should like to acknowledge the debt I owe to numerous friends and former teachers whose comments and criticisms have helped me in varying degrees to develop the ideas expounded in this thesis; among these have been my supervisor Dr. S. Blackburn, Mr. M. Dummett, Mr. Saul Kripke, and Mr. J. L. Mackie. I am particularly indebted to the interest and encouragement given me at various times by my good friend Dr. C. J. G. Wright. In addition I should like to thank my colleagues in the Department of Philosophy at the University of Birmingham for their support and consideration for my efforts to complete this thesis.

## Contents

|                                                                                        | page |
|----------------------------------------------------------------------------------------|------|
| Introduction                                                                           | 1    |
| Chapter I:   Frequencies and Individual Events                                         | 4    |
| Chapter II:   A Definition Applicable to Individual Events                             | 28   |
| Chapter III:   Randomness and Unique Probabilities                                     | 47   |
| Chapter IV:   Choosing The Narrowest Available Reference Class                         | 66   |
| Chapter V:   Two Concepts of Probability                                               | 117  |
| Chapter VI:   The Principle of Indifference and<br>The Classical Theory of Probability | 150  |
| Chapter VII:   The Principle and Its Critics                                           | 196  |
| Bibliography                                                                           | 254  |



## INTRODUCTION

In this thesis I will treat a number of problems in probability theory that seem to me of particular interest and which, I believe, have not as yet been satisfactorily resolved in the literature of the field. To begin with, I hope to formulate a definition of probability that will constitute a genuine frequency theory for the single case. While the possibility of such a theory has been considered for some time, much of the previous discussion of this problem has been unsystematic, inconclusive and, on occasion, inconsistent. In Chapter I I will consider some of the history surrounding this problem and in Chapter II I will present a theory which I believe represents the correct solution to it. Once properly formulated, a frequency theory for individual events can, I think, be shown to be preferable to the standard frequency theory in which probabilities are denied to individual events. In arguing for this I will, in later chapters, consider various issues in probability theory that are, in the first instance, of interest quite independently of the problem of single case probabilities. In Chapter III I will argue that despite Von Mises' explicit denial of probabilities to individual events, his requirement of randomness is best understood as a consequence of a frequency theory in which genuine probabilities are assigned to individual events. Besides supporting the line of argument of the earlier chapters, this discussion hopefully enhances our understanding of Von Mises' celebrated requirement.

The discussion in Chapter III will involve, crucially, a well-known adjunct to frequency theories of probability, the principle of choosing the narrowest reference class for which statistics are available. In Chapter IV I will examine this (and the related requirement of total evidence) at length for I believe there are a number of interesting, but in some cases hitherto unexplored, problems involved in stating this principle rigorously. In Chapter V I will turn to Carnap's famous thesis that there exist two distinct concepts of probability - one statistical, one epistemic - and will show how the frequency theory for the single case I have offered provides a means for reconciling these two concepts. My frequency theory for the single case will turn out (on the analysis given in Chapter V) to be one in which probability is defined as a quantitative and empirical relation between evidence and hypothesis; from this vantage point Carnap's two concepts of probability will appear to arise only from the existence of two distinct means for arriving at numerical measurements of the relation of evidential support. At the end of Chapter V I will discuss at some length the problem of subjectivism in theories of probability.

In the last two chapters of the thesis I will argue that the classical theory of probability taken with the famous, or if you prefer infamous, principle of indifference actually presented in a rudimentary and somewhat muddled way a definition of probability closely related to that found in Chapter V to underlie Carnap's two (apparently) distinct concepts of probability. In particular I believe the principle of indifference itself constitutes a semantic definition of probability - one in which probability is defined as a comparative relation between evidence

and hypothesis. This interpretation of the principle - which I believe to be novel - allows us, on the one hand to understand why the principle has so long fascinated writers on probability theory, for the definition of probability it encapsulates is highly plausible. On the other hand, this interpretation also allows us to understand why the principle has so often been criticised, for the criticisms levelled against it can be seen as natural, but not insurmountable, objections to the definition of probability encapsulated in the principle and the rudimentary manner in which that definition is expressed.

Given the myriad controversies that have arisen over the principle, the discussion of it in the last two chapters of this thesis is of necessity quite long. This, unfortunately, has meant that I have had to forego consideration of a number of issues relevant to various other parts of the thesis. Among the relevant topics that I have been unable to consider are the problem of induction and recent propensity theories of probability. Other questions receive only passing attention, where a full discussion would have been more desirable. By considering a limited number of problems at length, at the expense of other topics, I have hoped to gain in depth what has been lost in scope.

## CHAPTER I

### Frequencies and Individual Events

The history of the theory of probability is the history of attempts to find an explication for the prescientific concept of probability.

R. Carnap, Logical Foundations of Probability, p. 23.

One rather well known attempt at explicating the concept of probability is the definition of probability as relative frequency in the long run. There are numerous objections to this definition, only one of which I shall be concerned with here. It is the objection that probability, when defined as relative frequency in the long run, becomes a relational property not of individual events but of classes of events and that such a conclusion does not accord with our intuitions on the nature of probability. The aim of the first two chapters of this thesis is to develop a theory which, although defining probability in terms of relative frequency, does not do so in a manner that makes probability solely a relational property of classes of individual events.

I will take it for granted that some frequency definition of probability is desirable. Thus many of the usual objections to the frequency theory, except the one already mentioned, will be ignored in these first two chapters. That the frequency theory presupposes induction, that statements



about infinite sequences of empirical data are unverifiable, etc. are objections the merits of which will not concern me in these chapters. My basic position is that while the standard frequency theory may or may not meet these objections in some plausible fashion, it still would have to be rejected as an adequate definition of our prescientific concept of probability because it makes probability a property of, or relation between, classes of individual events. I will not argue at length that our prescientific concept of probability does genuinely apply to individual events, because the grounds for such an argument are only our intuitions on the concept of probability; hence what "argument" there is can be no more than an appeal to these intuitions.

It is perhaps best to begin with a straightforward exposition of the standard frequency theory. In point of fact the widespread appeal of the standard frequency theory has insured it numerous formulations; the following, combining features of Reichenbach's and Von Mises' formulations, seems most appropriate in view of the subsequent discussion of randomness. Given as an ordered pair the infinite class  $A$  of repeated experiments  $x_1, \dots, x_1, \dots$  and another infinite class  $B$  of experiments (possibly but not necessarily a sub-class of  $A$ ), probability is defined as a function that assigns to the ordered pair of classes  $\langle A, B \rangle$  a real number equaling the limit  $p$  of the relative frequency of  $B$  on  $A$ . The class  $A$  is traditionally called 'the reference class' - though 'collective' is also used if the further condition of randomness is fulfilled - while the class  $B$  is generally known as the attribute class.

As is customary this definition of probability consists in specifying a particular function which is known as 'a probability'. Throughout the thesis I will take the specification of a particular function as the standard method of defining a quantitative concept of probability. For convenience I will often then go on to adopt the common philosophical usage in which such a definition is referred to by the nature of the domain of the probability function specified i.e. in the definition given above probability is defined as a 2-place relation on classes of individual experiments.

This definition can be illustrated - and indeed made to appear quite plausible - by the simple example of a coin. Our repeated experiments  $x_1 \dots x_1 \dots$  consists of tossing the coin and examining to see if it is heads or tails. The class A consists of all and only tosses of this coin, the class B consists of tosses of coins which land heads\*; accordingly ' $x_1 \in A \cdot x_1 \in B$ ' signifies that

---

\* In this example, as well as in the precise statement of the frequency definition, the reference class A consists of all and only the elements of a given sequence of experiments. This parallels Von Mises' definition of probability in relation to a collective but is slightly less general than Reichenbach's definition. In Reichenbach, sequences of experiments (such as the one in our definition) containing all and only members of a particular reference class are said to be 'compact' on that reference class and, although these sequences figure prominently in Reichenbach's work, his definition of the probability of an attribute class on a reference class is formulated in terms of sequences which are not compact on that reference class.<sup>1</sup> One advantage of our less general definition is that we need not distinguish between reference classes of experiments and sequences of experiments. Technically a sequence of experiments will be a set of them under some order relation, e.g. order in time, while a reference class would be an unordered set of elements. If we are only concerned with reference classes on which some sequence is compact, we may speak indifferently of sequences of experiments or reference classes of experiments, for every reference class will contain all and only the elements of the sequence compact on it.

the first toss has landed heads, etc. Advocates of the standard frequency theory claim that defining a probability function which takes these classes as arguments and determines probability values in accordance with the limit of relative frequency of B on A corresponds to the ordinary manner of speaking in which we say the probability of 'heads' (B) with this coin (A) is its percentage in the long run. Admittedly there is always a degree of controversy over the identification of an unobservable limit of relative frequency with the ordinary idea of percentage in the long run, but except where I specifically deal with this question, I will accept the view of most frequentists that the limit of relative frequency in an infinite sequence is a plausible formulation of our ordinary concept of percentage in the long run.

To facilitate further discussion we should distinguish clearly between individual experiments  $x_1, x_2$  etc. which are possible members of the reference and attribute classes A and B and the event of such experiments actually belonging to those classes e.g.  $x_1 \in A$  or  $x_2 \in B$  etc. I will use the expression 'individual event' to designate the fact that a particular experiment occurs as a member of a reference class or attribute class. This terminology allows us to speak simply and naturally of the phenomena with which we are primarily concerned, the occurrence or non-occurrence of an individual experiment as a member of a given attribute class - this will be the occurrence or non-occurrence of an individual event. Moreover since every experiment in so far as it belongs to a reference class or attribute class constitutes an individual event in this sense, we may now adopt the customary description of the frequency



theory as a definition of probability as a 2-place relation on classes of individual events.

Before proceeding further we should note that it is possible to transform our definition of probability as 2-place relation on classes of individual events into an equivalent but simpler definition in terms of classes of sentences.<sup>2</sup> Here we simply exploit the correspondence - long noted in probability theory - between individual events and sentences asserting the occurrence of these events. Above probability was defined in terms of the limit of relative frequency between an infinite reference class  $A$  of individual experiments  $x_1, \dots, x_1, \dots$  and an attribute class  $B$ , to which the experiments might or might not also belong. If we form the sentential matrix ' $x_1 \in A$  and  $x_1 \in B$ ', we can generate an infinite class of sentences by substituting for  $i$ . To every experiment  $x_1$  of  $A$  that turns out also to be  $B$  will correspond a true sentence in the class formed by this matrix; to every experiment  $x_1$  of  $A$  that is not  $B$  will correspond a false sentence. Accordingly, if the relative frequency of  $B$  on  $A$  approaches a limit  $p$  as the number of experiments  $n \rightarrow \infty$ , then the percentage of true sentences formed by this matrix on the total number of sentences formed will also approach  $p$  as  $n \rightarrow \infty$ . If we call this percentage of true sentences among total sentences the truth frequency of the class of sentences, we may define probability as the limit of truth frequency of such a class of sentences. As for most purposes such a definition is equivalent to the one already given, I shall keep to the original formulation until much later in this thesis.



Probability defined as the limit of relative frequency between classes of events satisfies in a quite straight forward manner the commonly used mathematical formulation of the axioms of probability. However for our purposes we should instead consider the frequency theory's interpretation of the more traditional and less formal formulations of the axioms of probability. As the axiom systems found in non-mathematical writers on probability notoriously differ in detail, let us consider a simple axiom system containing the five axioms that figure prominently in almost every system.

$$A1 \quad (A \supset B) \supset P(A, B) = 1$$

$$A2 \quad (P(A, B) = p \cdot P(A, B) = q) \supset p = q$$

$$A3 \quad 0 \leq P(A, B) \leq 1$$

$$A4 \quad P(A, B \vee C) = P(A, B) + P(A, C) - P(A, B \cdot C)$$

$$A5 \quad P(A, B \cdot C) = P(A, B) \cdot P(A \cdot B, C)$$

Axiom 1, which with Axiom 3, is sometimes known as an axiom of normalization, asserts that if B is certain given A, then the probability of B on A is 1. Axiom 3 asserts that the probability function is non-negative and has values in the interval 0 - 1. Axiom 2, the axiom of univocality, asserts the uniqueness of probability values. Axioms 4 and 5 are, respectively, the well-known general addition and multiplication theorems.

On the standard frequency interpretation the symbols 'A', 'B', 'C', etc. are variables taking as values infinite classes of experiments. The operator 'P( )' is interpreted as a 2-place function on these classes, with

numerical values determined on the basis of limits of relative frequency.

Logical symbols within the scope of the probability operator are interpreted in terms of the set theoretic equivalents of ordinary logical operations, e.g.

$P(A, B \vee C)$  is the probability of the union of B and C on A and  $P(A, B \cdot C)$  is that of the intersection of B and C on A. To each axiom corresponds a

mathematical truth concerning limits of relative frequency, e.g. trivially for

the case of axiom 2 we have the truth that there is at most one limit of relative frequency for an attribute in a given infinite sequence. Axioms 4 and 5 are

obviously more interesting and it will be instructive if we consider the frequency interpretation of one of them. On the frequency theory's interpretation Axiom 5

states that the limit of relative frequency of the class  $B \cap C$  on A equals the

product of the limit of B on A and the limit of C on  $A \cap B$ . Proof, as Reichenbach points out<sup>3</sup>, is virtually immediate. Using Reichenbach's notation we may

abbreviate the relative frequency of  $\psi$  on  $\phi$  in n trials as  $F^n(\phi, \psi)$  and the

number of  $\phi$ 's that have occurred in n trials as  $N^n(\phi)$ . It is clear that the

relative frequency in n trials of  $\psi$  on  $\phi$  is  $\frac{N^n(\phi, \psi)}{N^n(\phi)}$  and so we obtain:

$$\begin{aligned} F^n(A, B \cdot C) &= \frac{N^n(A, B \cdot C)}{N^n(A)} = \frac{N^n(A, B)}{N^n(A)} \cdot \frac{N^n(A, B \cdot C)}{N^n(A \cdot B)} \\ &= F^n(A, B) \cdot F^n(A \cdot B, C) \end{aligned} \quad (5)$$

The equalities in Reichenbach's equation (5) remain valid for transition to the limit and thus we have

$$\lim_{n \rightarrow \infty} F^n(A, B \cdot C) = \lim_{n \rightarrow \infty} F^n(A, B) \cdot \lim_{n \rightarrow \infty} F^n(A \cdot B, C),$$

which in the frequency interpretation is axiom 5.

The intuitive meaning of axiom 5 as interpreted by the frequency theory perhaps comes out more clearly in a simple schema used by Reichenbach to represent occurrences of events in  $A$  that are and are not members of the attribute classes  $B$  and  $C$ :

|                 |           |           |                 |           |           |           |           |     |
|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------|-----------|-----|
| $A$             | $A$       | $A$       | $A$             | $A$       | $A$       | $A$       | $A$       |     |
| $B$             | $\bar{B}$ | $\bar{B}$ | $B$             | $\bar{B}$ | $B$       | $\bar{B}$ | $B$       | (7) |
| $\underline{C}$ | $C$       | $\bar{C}$ | $\underline{C}$ | $C$       | $\bar{C}$ | $C$       | $\bar{C}$ |     |

With it understood that the frequency he is concerned with is always relative frequency on the reference class  $A$ , Reichenbach explains:

The frequency  $F^n(A, B, C)$  represents the frequency of the couples  $B, C$ ; the first of the expressions standing on the right side of (5),  $F^n(A, B)$ , counts the frequency of  $B$ . Now  $B$  selects from the sequence of  $C$ 's a subsequence, the elements of which are marked by a lower double bar in (7); this subsequence, of course, contains elements  $C$  as well as  $\bar{C}$  [i.e. events not in the class  $C$ ]. The number of elements of this subsequence is given by  $N^n(A, B)$ ; therefore  $F^n(A, B, C)$  means the relative frequency of  $C$  in the subsequence... Formula (5) states that the desired frequency of the pair  $B, C$  can be represented as the product of that frequency  $B$  by the frequency of  $C$  counted within the subsequence selected by  $B$ .

Having looked at the details of the frequency theory's definition of probability - and in particular the interpretation it provides for the axioms of probability - we may now turn to the main concern of this chapter, the question of determining probabilities for individual events. As should be abundantly clear by now the standard frequency theory defines a concept of

probability in which, strictly speaking, individual events are not assigned probabilities. Probabilities are only assigned to classes of individual events, or rather pairs of such classes. The obvious conflict between such a view and our ordinary discourse in which we habitually assign probabilities to individual events (and indeed do so in light of observed frequencies) has pressed advocates of the frequency theory to mount a spirited, if often vacillating, defense of this central feature of their theory. The main (and indeed usually the only) argument put forward by them is the well known objection that the assignment of probabilities to individual events on the basis of class frequencies will yield conflicting results depending on the class chosen. That is to say if we stipulate, as seems natural, that the probability of an individual experiment being the member of a given attribute class B is to be measured by the relative frequency of B on some infinite reference class, we must acknowledge that we will determine different probabilities for the individual event if B displays different frequencies on different reference classes. Von Mises, for example, produces several instances of this difficulty and then goes on to conclude: 'We shall not speak of probability until a collective has been defined'. It is clear from the context that he believed this to rule out the assignment of probabilities to individual events.<sup>4</sup>

However, immediately after having announced this position, contradictions begin to appear in the very examples he uses to illustrate it. This is perhaps clearest with a celebrated example he borrows from Laplace,



that of our surprise on obtaining the word 'Constantinople' in a random draw from cards containing single letters of the alphabet. The interest of this example is that all combinations of 14 letters are equally improbable so, on first analysis, the appearance of the word 'Constantinople' should not be any more surprising than that of any other combination of letters. Von Mises' analysis is that in the collective of repeated draws of 14 letters, combinations having the attribute 'meaningless' will vastly predominate over those that are 'meaningful'; accordingly the appearance of the attribute 'meaningful' in a given case is exceedingly improbable.

This analysis seems correct up to a point and does, indeed, underline Von Mises' contention that a collective and an attribute are central to ascriptions of probability; however, it is perfectly clear that this example concerns the probability (or improbability) of an individual event. It is the appearance of the attribute 'meaningful' (under the guise of 'Constantinople') on a particular draw which is deemed to have been an extremely improbable event. The conflict here between Von Mises' claim that individual events cannot have probabilities and his reliance on such probabilities in his examples is just one instance of the vacillating position which advocates of the standard frequency theory take up over individual events.

The difficulties inherent in their explicit denial of probabilities to individual events becomes more and more apparent when they attempt to give some account of the role of probability in guiding our actions in the face of uncertainty over an individual event. Reichenbach, for one, readily admits

that practical exigencies demand an account of how long run frequencies relate to behaviour in the face of individual events whose outcome is uncertain; in fact he begins and ends his treatment of the single case with particular emphasis on the need for an account of the single case that does justice to our actions in the face of uncertainty<sup>5</sup>. His solution is ultimately in terms of 'weighted posits' but the essence of his argument can be stated quite clearly without reference to his concept of posits:

If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the problem of the reference class. Assume that a case of illness can be characterized by its inclusion in the class of cases of tuberculosis. If additional information is obtained from an X-ray, the same case may be incorporated in the class of serious cases of tuberculosis. Depending on the classification, different probabilities will result for the prospective issue of the illness.

We then proceed by considering the narrowest class for which reliable statistics can be compiled. If we are confronted by two overlapping classes, we shall choose their common class.

... We do not affirm that this method is perfectly unambiguous. Sometimes it may be questioned whether a transition to a narrower class is advisable, because, perhaps, the statistical knowledge regarding the class is incomplete. We are dealing here with a method of technical statistics; the decision for a certain reference class will depend on balancing the importance of the prediction against the reliability available. It is no objection to this interpretation that it makes the probability constructed for the single case dependent on the state of our knowledge. This knowledge may even be of such a kind that it does not determine one class as the best. ...

Whereas the probability of a single case is thus made dependent on our state of knowledge, this consequence does not hold for a probability referred to classes. If the reference class is stated, the probability of an attribute is objectively determined, though we may be mistaken in the numerical value we assume for it on the basis of inductions. ... The dependence of a single-case probability on our state of knowledge originates from the impossibility of giving this concept an independent interpretation; there exist only substitutes for it, given by class probabilities, and the choice of the substitute depends on our state of knowledge. My thesis that there exists only one concept of probability, which applies both to classes and to single cases, must therefore be given the more precise formulation: there exists only one legitimate concept of probability, which refers to classes, and the pseudoconcept of a probability of a single case must be replaced by a substitute constructed in terms of class probabilities. ...

I regard the statement about the probability of the single case, not as having a meaning of its own, but as representing an elliptical mode of speech. In order to acquire meaning the statement must be translated into a statement about a frequency in a sequence of repeated occurrences. The statement concerning the probability of the single case thus is given a fictitious meaning, constructed by a transfer of meaning from the general to the particular case. The adaptation of the fictitious meaning is justifiable, not for cognitive reasons, but because it serves the purpose of action to deal with such statements as meaningful.

Reichenbach completes his account by adding that the numerical value of the frequency in the reference class eventually chosen for an individual event is the 'weight' to be assigned to the proposition asserting the occurrence of that event by our 'fictitious transfer' from the general to the particular. The weight assigned to a proposition about an individual event determines our action

about that event, indeed to the point of determining the fair betting ratio for it. I will shortly examine the concept of weight in more detail as, later, I will consider the details of Reichenbach's criterion of choosing the narrowest available reference class; for the moment however, we should note the most salient feature of Reichenbach's account - that the concept of a probability of a single case is fictitious, indeed a 'pseudoconcept' for which "there exists only substitutes ... given by class probabilities".

Admittedly by introducing the concept of weight, or fictitious probability, of a single case determined by the narrowest available reference class, Reichenbach (as we will see in more detail later) provides much of what is necessary for an account of how long-run frequencies can be used in determining our action over single cases. But equally clearly Reichenbach has not given a theory in which genuine probabilities are assigned to individual events, such probabilities are for him fictitious and the resulting concept of probability is a pseudoconcept, whatever that may be. If a definition of probability were given which genuinely applied to the single case, the resulting probabilities could not be fictitious; or, to put it another way, if the semantics of a concept have been properly given, the meanings so given are in no way fictitious. To characterize the meaning given to the concept of a single case probability as fictitious is simply to admit that such a concept has not been properly defined.

This point emerges clearly in the recent writings of Salmon, a lucid follower of Reichenbach. Salmon remarks:



Reichenbach admits that the meaning of "probability" for the single case is fictitious. He does not offer rules which assign probability values univocally. It is apparent that very practical considerations determine the value actually assigned. It would have been better, I think, if he had refused to apply the term to single events at all but had reserved some other term such as weight which he often used for this purpose. We could then say that probability is literally and exclusively a concept that applies to infinite sequences, not to single cases.<sup>6</sup>

Thus despite Reichenbach's efforts we are back where we were before - without definition of probability which does justice to our prescientific intuitions that probability does apply to individual events.

Salmon himself, however, offers a new version of Reichenbach's proposal for treating the single case along frequentist lines. Salmon first proposes a procedure for selecting the correct reference class for a single case which is essentially equivalent to Reichenbach's selection of the narrowest available reference class; Salmon's formulation (to be studied in detail later) is in terms of selecting the broadest homogeneous reference class. In order to explain the role of this principle Salmon cites a distinction made by Carnap:

Carnap has made an extremely important and useful distinction between inductive logic and the methodology of induction... Inductive logic contains the systematic explication of degree of confirmation, and analytic degree of confirmation statements that hold for a given confirmation function in a given language. The methodology of induction contains rules for the application of inductive logic - that is, rules that tell us how to make use of the statements of degree of confirmation in deciding courses of practical action. As I indicated the requirement of total evidence is one of the important methodological rules, and the rule of maximising estimated utility is another.<sup>7</sup>

Salmon then employs this distinction to clarify Reichenbach's original proposal and we should consider Salmon's argument in full, even if it involves some repetition from above:

Reichenbach, unfortunately, did not make the same clear distinction between probability theory proper and the practical rules for the application thereof. Such a distinction would have been helpful particularly for the problem of the single case.

Reichenbach admits that the meaning of "probability" for the single case is fictitious. He does not offer rules which assign probability values univocally. It is apparent that very practical considerations determine the value actually assigned. It would have been better, I think, if he had refused to apply the term to single events at all but had reserved some other term such as weight which he often used for his purpose. We could then say that probability is literally and exclusively a concept that applies to infinite sequences, not to single cases. If we want to find out how to behave regarding a single case, we must use probability knowledge; the problem is one of deciding how to apply such knowledge to single events. Our rules of application would tell us to find an appropriate reference class to which our single case belongs and use the value of the probability in that infinite sequence as the weight attached to the single case. The rule of selecting the broadest homogeneous reference class becomes a rule of application, not a rule for establishing values of probabilities within the theory proper... This treatment of the single case shows a deep analogy between the requirement of total evidence and the rule for selecting the appropriate reference class. Both are rules of methodology not of probability theory. Each rule requires us to make use of all relevant available information in arriving at practical decisions.<sup>8</sup>

The problem with Salmon's proposal is that it may be well and good to distinguish the definition of probability offered by a particular theory from the details of applying that definition in actual circumstances with a particular aim in mind, but the definition of probability offered by Salmon (and Reichenbach) can not be applied to single cases at all. In their theories the only genuine arguments to the probability function are classes of individual events and so no 'application' of that function can be made to individual events.

This difficulty is highlighted by Salmon's own words for he tells us at one point that 'probability is literally and exclusively a concept that applies to infinite sequences, not single cases' and yet it is rules of application for, presumably, the concept of probability which are intended to deal with the single case.

Thus despite his comments on their similarity, Salmon's proposal differs considerably from Carnap's original suggestion. Carnap originally put forward the requirement of total evidence as a principle for selecting in any given situation one particular evidence sentence  $e$  that should be used as one argument for the two place confirmation function he defined; this in turn yields a unique degree of confirmation for our hypothesis  $h$  which is the other argument to the confirmation function. With Salmon's proposal we are, in a given situation, able to choose one particular reference class: however treating this as an argument for the probability function defined by Salmon does not yield a probability value for the individual event we are concerned with, for the only other arguments to Salmon's probability function are classes of events. The

point here is that Carnap's methodological rules give application to his definition of confirmation in that they tell us which among the many confirmation values determined by his (relational) theory of confirmation is the one most appropriate for a given situation. But when faced with an individual event, any probability values determined by Salmon's theory are irrelevant for they are all probabilities of classes; accordingly methodological rules for applying that definition do not help in the single case.

Now, of course, in pointing this out, I am deliberately denying Salmon's proposal a certain grain of salt, for his intentions are clear enough, even if they are somewhat muddled. However that Salmon is in a genuine muddle is important for us as it is just another instance of the vacillation mentioned earlier among frequentists over the single case. In Von Mises single case probabilities were ostensibly banished only to reappear in examples like that of Laplace. In Reichenbach's single case probabilities are tentatively countenanced and then treated as fictitious. Salmon, following on, quite rightly corrects Reichenbach's infelicity in admitting such dubious entities as fictitious probabilities - only to find himself treating the single case as an alleged application of a theory that admittedly does not apply to such cases.

It is obvious by now, I should think, that the difficulty which advocates of the standard frequency theory have over the single case springs from their insistence that the only genuine concept of probability applies to class<sub>es</sub> of individual events. Not surprisingly with this assumption in the background, any attempt at providing a theory which does apply to individual events must contain some vacillation or inconsistency. A more promising



approach would be to abandon the contention that the only legitimate concept of probability is that of class frequencies and, rather, adopt the view that individual events have genuine probabilities in virtue of these class frequencies. The constructive theory put forward in the next chapter is an attempt to do this in a systematic and perspicuous way. Before presenting that theory, it will be worth our while to consider anticipations of it in the work of two eminent writers on probability, Keynes and Popper. Both have proposed treatments of the single which closely resemble those of Salmon and Reichenbach but which, promisingly, do (almost) unequivocally permit genuine probabilities for the single case.

Popper, for example, after commenting that the frequency theory usually rules out statements ascribing probabilities to the single case remarks:

... It is easy, however, to interpret these statements as correct, by appropriately defining formally singular probabilities with the help of the concept of objective probability or relative frequency. I use ' ${}_αP_k(β)$ ' to denote the formally singular probability that a certain occurrence  $k$  has the property  $β$ , in its capacity as an element of a sequence  $α$  - in symbols  $k ∈ α$  - and I then define the formally singular probability as follows:

$${}_αP_k(β) = {}_αF(β) \quad (k ∈ α) \text{ (Definition)}$$

This can be expressed in words as: The formally singular probability that the event  $k$  has the property  $β$  - given that  $k$  is an element of the sequence  $α$  - is, by definition, equal to probability of the property  $β$  within this reference sequence  $α$  ...

As the definition shows, a formally singular probability statement would be incomplete if it did not explicitly state a reference-class. But although  $α$  is often not explicitly mentioned, we usually know in such cases which  $α$  is meant. ...

In many cases there may be several different reference sequences for an event  $k$ . In these cases it may be only too obvious that different formally singular probability statements can be made about the same event. ... It is not possible to lay down a general rule as to which out of several possible reference-classes should be chosen. (The narrowest reference-class may often be the most suitable, provided that it is numerous enough to allow the probability estimate to be based upon reasonable statistical extrapolation, and to be supported by a sufficient amount of corroborating evidence.)

Obviously this proposal, down to a qualified recommendation of the principle of choosing the narrowest reference class, closely parallels that found in Reichenbach. Probabilities are still defined in the first instance as properties of sequences and only derivatively assigned to individual elements. But while this leads us to two apparently distinct concepts of probability - one applying to sequences, the other to individual elements - at least Popper does not regard the derived probabilities as fictitious.

Keynes's proposal, credited in part to Whitehead, while quite similar to Popper's and Reichenbach's, even more clearly endorses the validity of probabilities assigned to individual events in virtue class frequencies. As with Popper the probability assigned to an individual event (or rather a proposition about an individual event), is, for Keynes, to be measured by the frequency of true propositions in a class of similar propositions. Again as in Popper, the probability for a single case will vary depending on the class chosen, for the proposition about the individual event will belong to different classes with different frequencies of truth. However, unlike Popper, the probability of the single case is for Keynes the uniquely correct concept of probability - the class

frequencies which determine the probabilities for the single case are just that, frequencies and not probabilities.<sup>10</sup>

While these proposals of Popper and Keynes are intuitively attractive, they are, alas, little more than that. Their proposals simply consist in stating the practical precept that we are to measure the probability of an individual event (or proposition about it) by reference to class frequencies. How far this is from a full theory of probability is apparent when we realize that the very minimum condition for an acceptable theory of probability is that it specify a probability function which is an adequate semantic interpretation of the axioms of probability. Clearly the precept advocated by Keynes and Popper does not, in unelaborated form, provide a semantic interpretation of the axioms of probability; nor, for that matter, do Keynes or Popper clearly indicate any one probability function determined by their theories. For example in Popper it is unclear whether probability is being defined by one or two functions - if one, how can legitimate probabilities apply both to classes of events and individual events. If two - one for classes of events and one for individual events - what is the relation between them? Are we to take both as independently acceptable interpretations of the axiomatic concept of probability or is one somehow dependent on the other?

In Keynes the matter is even worse for his frequency theory for the single case makes the probability of the single case relative to a reference chosen and the data available at the time we make that choice.<sup>11</sup> Thus at times it looks as if the probability function Keynes has in mind is 2-place function on

propositions and classes to which those propositions belong, while at other times it looks as if it is a 2-place function between propositions and data about those propositions. He clearly intends the latter when he investigates the adequacy of the theory as an interpretation of the axioms of probability, but (not surprisingly in light of the ambiguity over the arguments to the probability function) he eventually concludes that this theory does not provide an adequate model for the axioms of probability and so should be rejected.<sup>12</sup>

To his credit, though, Keynes, unlike Popper, at least investigated the adequacy of a frequency theory for the single case as an interpretation of the axioms. Moreover, despite the fact that proposals like Keynes's to determine probabilities for the single case by reference to class frequencies have had a certain currency in more recent literature, no attempt to my knowledge has been made to elaborate any of them to the very minimal level of providing an adequate interpretation of the axioms. The difficulties involved here become most apparent when we think of the axioms in their non-mathematical formulation, given above in five axioms.

Let us assume that proposals along the lines of Keynes and Popper do provide a means for determining the probability of an individual event  $B$  relative to some constant  $A$  (whatever, in fact,  $A$  will turn out to be on such a proposal); the problem is that it still is not clear how these proposals envision determining probabilities for individual events  $B$  and  $C$  relative to  $A$ , or  $C$  relative to  $A$  and  $B$ , for that matter. Without



explanations of this kind we can not begin to investigate, much less prove that, for example, Axioms 4 and 5 come out valid.

In point of fact Axiom 2 presents the most egregious difficulties for proposals along the line of Keynes and Popper. Axiom 2 states that probability values are to be unique and both Keynes and Popper admit that their proposals permit numerous different probabilities for a given single case. Moreover, this variability of single case probabilities was, of course, the basic reason Von Mises, Reichenbach and Salmon denied genuine probabilities to individual events.

Now to my knowledge neither Von Mises, Reichenbach nor Salmon ever object to the variability of probabilities for the single case explicitly on the grounds that such variability violates the axiom of univocality. But whatever other difficulties may be raised by the variability of single case probabilities e.g. for the purposes of action, the first and most basic problem is that raised by the axiom of univocality, for, as already indicated, the minimum requirement for a satisfactory definition or semantics for the concept of probability is that it specify a function which is adequate as an interpretation of the axioms of probability.

In point of fact the variability of single case probabilities does not present an insurmountable difficulty in constructing the semantics of a frequency concept of probability applicable to the single case; indeed the apparently troublesome variability itself suggests the probability function

which will provide an adequate 'frequency' interpretation of the axioms. The details of this semantics will be presented in the next chapter; we may summarize this chapter by noting again why the proposals in Reichenbach, Salmon, Keynes and Popper for single case probabilities do not, as they stand, provide the required semantics. Reichenbach and Salmon, as we saw in detail, specify probability functions which take only classes of events as arguments and this point alone ensures that their treatment of the single case can hardly be adequate. Keynes and Popper avoid this difficulty but their proposals are too rudimentary to fulfill the minimum adequacy condition for a definition of probability: neither clearly specifies one definite probability function, nor are their proposals shown to provide an adequate model for the axioms of probability.

## Footnotes for Chapter I.

1. Reichenbach, H., The Theory of Probability, Berkeley and Los Angeles, 1949 pp. 68-69.
2. Ibid.,
3. Ibid., pp. 74 ff
4. Von Mises, R., Probability Statistics and Truth, Second edition London 1957, p. 18.
5. Reichenbach, op. cit., Sections 72 and 73.
6. Salmon, W.C., The Foundations of Scientific Inference, Pittsburgh, 1967 p. 93.
7. Ibid.,
8. Ibid., pp. 93-94.
9. Popper, K., The Logic of Scientific Discovery, revised ed., London 1968 pp. 202-204.
10. Keynes, J.N., A Treatise on Probability, London 1921, pp. 100 ff.
11. Ibid., p. 102.
12. Ibid., pp. 105-107.

## CHAPTER II

### A Definition Applicable to Individual Events

At the close of the previous chapter, it was suggested that the variability of single case probabilities could lead to a 'frequency' interpretation of the axioms of probability genuinely applicable to individual events. Now since the probability of an individual event varies with the reference class chosen, the probability of an individual event is, formally speaking, a relation between it and a reference class. Now in fact it is widely accepted that the most natural axiom systems for probability are relational in character, that is, employ a 2-term probability operator. In the informal system given above, the five axioms all embody a two-term probability operator and very similar formulations have been put forward by Keynes, Kneale, Carnap, Popper, Reichenbach and Salmon, to name just a few. In the mathematical formulation of the axioms the relational character of probability is less overt, but undeniable. Probability is defined in terms of a measure function on sub-sets of an original sample space  $S$ . All mathematical treatments emphasize that a probability function can only be defined relative to such a sample space and so, although overtly a one-term function, the probability function can be thought of as having tacitly a second argument, namely the sample space  $S$  presupposed.



Since the variable probability of an individual event can be thought of as a relation between it and a reference class, we can, to put it crudely, avail ourselves of this relation in giving a semantic interpretation for the similarly relational axioms of probability. To put this more exactly let us return to the symbolism used earlier in the previous chapter in explaining the standard frequency theory's definition of probability as a 2 place function on classes of individual events. There ' $x_n \in A$ ' represented the fact that the  $n^{\text{th}}$  experiment in a sequence belonged to a reference class A, ' $x_n \in B$ ' represented a favourable outcome of that experiment, namely that it was a member of the attribute class B. Now in considering the problem of assigning probability to this particular outcome on the basis of the long run frequency of B, we noted that the probability of the individual event  $x_n \in B$  will vary as the attribute B approaches different limiting values of relative frequency on different reference classes A; that is, the probability of  $x_n \in B$  varies as the experiment  $x_n$  is thought of as belonging to different reference classes. This suggests that we define probability as a 2 place function that assigns real numbers in the interval 0 - 1 to ordered pairs consisting of the individual event of an experiment belonging to a particular reference class ( $x_n \in A$ ) and the individual event of that experiment belonging to the attribute class we are interested in ( $x_n \in B$ ). The numerical degree of probability assigned to these ordered pairs of individual events is to be determined by the limiting value of the relative frequency of the attribute class on the reference class. In contrast to the standard frequency definition of probability as a function from a set of ordered pairs of reference class and attribute class, probability here is defined as a function from a set of ordered pairs, each pair consisting, respectively,

of the individual event of a given experiment belonging to a particular reference class and the individual event of that experiment belonging to an attribute class.

Put more intuitively this definition of probability comes to the following: we are interested in the probability of the occurrence of a certain individual event, the favourable outcome of a given experiment - this can be thought of as the probability of that experiment belonging to a particular attribute class. The probability of this event's occurrence will vary depending on which of the different conditions under which the experiment is held is considered - the different reference classes to which the experiment belongs. As the probability of the favourable outcome so varies with the reference class the experiment is thought to belong to, the probability of that outcome is defined only in relation to another individual event consisting of the experiment belonging to a particular reference class. Thus the elements in the domain of our probability function are ordered pairs consisting of the individual events of an outcome of an experiment coupled with the event of the experiment belonging to a certain reference class. In order to get a frequency definition of probability we define our probability function from the set of such ordered pairs of individual events so that the numerical values assigned as the degree of probability to these pairs are determined by limiting values of relative frequency of attribute classes on reference classes. By having our probability function assign the limiting value of relative frequency between attribute and reference class as probability value to the ordered pair of individual events, we give a definition of probability that formally expresses the intuition that individual events have genuine probabilities but only in virtue of long run class frequencies.

In accordance with philosophical usage this definition, like that of the standard frequency theory, will often be referred to by the nature of the domain of the probability function i.e. as a definition of probability as a (2 place) relation on individual events. Such usage highlights the central feature of the definition proposed here, namely that genuine probabilities are assigned to individual events, such as outcomes of experiments, but only relative to another individual event representing conditions under which the experiment is held.

This definition of probability entails that a statement of probability about an individual event  $x_n \in B$  will be incomplete unless it includes mention of antecedent event  $x_n \in A$ . This is not to say, however, that ordinary usage necessarily contains reference to such 'antecedent' individual events. Indeed, ordinary usage, on the above analysis, must be construed as elliptical, as ordinary probability statements about individual events  $x_n \in B$  do not always, or even often, contain reference to the additional term  $x_n \in A$  indicated here as necessary.

Construing ordinary probability statements as elliptical for relational statements is a widely accepted precept in various theories of probability; such an analysis is most often found in theories which make probability a relation between evidence and hypothesis. In general the evidence tacitly relied on but not explicitly referred to is taken to be the total available relevant evidence. Although I will discuss the role played by the total available evidence in much detail later, we should note for the

moment that the principle of selecting the narrowest available reference class corresponds closely to employing the total available relevant evidence. (A similar point was made, as we saw, by Salmon in terms of choosing the broadest homogenous reference class). This naturally suggests that in analyzing probability statements in ordinary usage which do mention a specific reference class, we can understand the speaker to be relying on the narrowest available reference class. For example, if, as is done in the United States, a numerical probability is assigned to the chance of rain for a given day, we should understand that this value is the expected limit of relative frequency of rain in the class of days most similar to the one in question and for which adequate statistics are available, i.e. the narrowest reference class for which statistics are available.

Having seen how our relational definition of probability for the single case accords with apparently non-relational probability statements, we must now consider its adequacy as a semantic interpretation of the axioms of probability. This is a relatively straightforward matter once the relational character of these probabilities is recognized. In fact the demonstration of the adequacy of this interpretation follows closely, but not exactly, that already given for the standard frequency theory. Considering our system of five axioms, we may recall that the standard frequency definition interpreted the symbols 'A', 'B', 'C' etc. as variables taking as values infinite classes of experiments. For our frequency theory for the single case these symbols are interpreted as variables whose values are individual events consisting of a particular experiment  $x_n$  belonging to just these classes i.e.  $x_n \in A$ ,  $x_n \in B$ , etc.



The probability operator ' $P( \quad )$ ' was interpreted in the standard frequency theory as a 2-place function on infinite classes of experiments; it is now interpreted as a 2-place function on individual events involving these experiments' membership in the infinite classes. The numerical values of probability for the function are determined by the limit of relative frequency between the classes to which the particular experiments belong i.e.  $P(A, B)$  is interpreted as the probability of  $x_n \in B$  given that  $x_n \in A$  and equals the limit of  $B$  on  $A$ . To keep this interpretation of the probability operator in mind I will use a small 'p' to indicate when it is intended i.e.  $p(x_n \in A, x_n \in B)$  is to represent the probability of the event  $x_n \in B$  in relation  $x_n \in A$  and equals the limit of  $B$  on  $A$ .

In this interpretation logical symbols are throughout given their ordinary meaning; however, when these symbols occur within the scope of a probability operator they are obviously equivalent to the set-theoretic operations employed in the standard frequency interpretation. For example,  $P(A, B \vee C)$  is interpreted as  $p(x_n \in A, x_n \in B \text{ or } x_n \in C)$ , which is obviously equivalent to  $p(x_n \in A, x_n \in B \cup C)$ . This, of course, equals the limit of relative frequency of  $B \cup C$  on  $A$ .

Since the values of the probability function are all determined by limits of relative frequency, the axioms and theorems of the probability calculus

come out as valid under this interpretation in exactly the same fashion they did under the standard frequency theory's interpretation. Two examples should suffice to illustrate this:  $P(A, B)$ , the probability  $p(x_n \in A, x_n \in B)$ , equals the limit of relative frequency of B on A. As only one limit of relative frequency can exist for an attribute class in a given sequence A, Axiom 2 of univocality comes out as valid, and, of course, it is in virtue of the uniqueness of the limit of relative frequency that the standard frequency theory satisfied Axiom 2. What doubts may have existed before over the capacity of a frequency theory for the single case to satisfy the axiom of univocality are now seen to be groundless. Axiomatically probability is a binary operator and thus the interpretation given to it is to be a 2-place function. The probability function on individual events we have defined is indeed a 2-place function corresponding to the variable or relational probability of a given individual event. The axiom of univocality only requires that between any two arguments to the function there exist at most one probability and, as we have seen, this is insured by our definition of probability.

The general multiplication theorem also provides a clear illustration of the adequacy of our definition of probability as an interpretation of the probability calculus. Axiom 5 states that

$$P(A, B \cdot C) = P(A, B) \cdot P(A, B, C)$$

In our theory this comes out as:

$$(1) \quad p(x_n \in A, x_n \in B \text{ and } x_n \in C) = p(x_n \in A, x_n \in B) \cdot p(x_n \in A \text{ and } x_n \in B, x_n \in C)$$

This in turn is equivalent to

$$(2) \quad p(x_n \in A, x_n \in B \cap C) = p(x_n \in A, x_n \in B) \cdot p(x_n \in A \cap B, x_n \in C)$$

As the values of the probability function  $p(x, y)$  are determined by limits of relative frequency of  $y$  on  $x$ , (2), in the symbolism found in chapter I, is

$$\text{equivalent to } \lim_{n \rightarrow \infty} F^n(A, B \cap C) = \lim_{n \rightarrow \infty} F^n(A, B) \cdot \lim_{n \rightarrow \infty} F^n(A \cap B, C)$$

This is exactly the same theorem employed in the standard frequency theory's interpretation of axiom 5, the proof of which we have considered already.\*

Having seen the adequacy of our theory as an interpretation of the axioms of probability, we should now consider some philosophical issues raised by it. An important feature of our definition is, of course, the variability of the probability of an individual event. We do not try to avoid this frequently cited variability but rather understand it as stemming from the relational character of the concept of probability. Thus the theory provides us with a viable alternative to the skeptical conclusions of Von Mises, Reichenbach

---

\*For those who may wonder how this theory so readily supplies an adequate interpretation of the axioms - while a similar theory was found by Keynes not to do so - the answer is simple. Keynes, as indicated above, was unnecessarily vague about the domain of the probability function he intended. In considering his theory's interpretation of the axioms he did not take probability to be a relation between two propositions about an individual event, one mentioning a reference class. Had he done so - or even made probability a relation between a proposition about an individual event and a reference class to which the proposition belonged - his theory would have been sufficiently similar to the one offered here to supply an adequate interpretation of the axioms. Rather he made probability a relation between a proposition and data on which a reference class could be chosen and for reasons that can be found in his original text, this does not give a satisfactory interpretation to the axioms.

and Salmon already cited. Reichenbach, for one, claimed that the variability of the probability of an individual event on different reference classes indicated 'that there exists only one legitimate concept of probability which refers to classes' and that the probability of individual events was 'a pseudoconcept' that could at best be given a certain 'fictitious' meaning for the purposes of action. But the dependence of the probability of individual events on the reference class chosen does not indicate that such probabilities are illegitimate; rather it only indicates that the probability of an individual event is a relation between it and the reference class chosen, or more exactly, a relation between it and another individual event consisting of the experiment in question belonging to the reference class chosen.

In addition we may also note that contrary to his claim our relational probabilities for individual events are as 'objectively determined' as the class probabilities Reichenbach speaks of. The degree of probability between two individual events  $x_n \in A$  and  $x_n \in B$  is determined quite objectively by the limit of relative frequency of B in A - for example the probability of heads on a particular toss ( $x_n \in B$ ), given that it is a toss of a regular and symmetrical coin ( $x_n \in A$ ), is the limit of relative frequency of heads (B) in repeated tosses of regular and symmetrical coins (A).

We could, of course, make similar comments on Salmon's rejection of probabilities for individual events - in favour of 'weights' - as his reasons are essentially those of Reichenbach. Although Von Mises obviously advocated a similar line in his dismissal of probabilities for individual events,



the emphasis is somewhat different in his writings. He repeatedly argues that probabilities can only be spoken of in reference to a collective or, as we saw him put it, 'We shall not speak of probability until a collective has been defined'. Although, as we remarked, it seemed clear Von Mises believed this conclusion ruled out probabilities for individual events, we can now see that this conclusion does not actually do so. In the theory given above we indeed can not speak of a probability of an individual event until a reference class, or collective, has been defined, for the probability of such an event is a relation between it and another individual event consisting of the experiment in question belonging to a particular reference class. This of course explains how Von Mises could use examples concerning individual events to illustrate his dictum that probability can only be spoken of once a collective has been defined. In the example we considered it was indeed the low frequency of an attribute - 'meaningful' - in a collective that accounted for our surprise at the appearance of 'Constantinople', but it was the appearance of that attribute on a particular occasion that we found improbable. In making probability a relation between an individual experiment's membership in an attribute class and its membership in a reference class we see how probability (and improbability) can be legitimately ascribed to the attribute's appearance in a particular experiment, still maintaining that a reference class is essential to the probability statement.

In a like fashion we may now understand Reichenbach's assertion that probability statements about the single case are an 'elliptical mode of speech', requiring 'translation into a statement about frequency in a sequence'. Probability

statements about an individual event are indeed elliptical for statements involving mention of a sequence or reference class on which frequencies are displayed. However, elliptical modes of speech are not to be 'translated' into other idioms, they are to be expanded within their own. Thus, as we saw, any probability statement about an individual event which did not mention a reference class would have to be treated as elliptical and expanded to include such mention. But once in the properly expanded form our probability statements are not, as Reichenbach claims, solely about relations between classes of events, for individual events are given genuine, albeit relational, probabilities.

More generally we are now able to understand why advocates of the standard frequency theory could only give vacillating and contradictory accounts of the single case. In Von Mises we found examples which contradicted his stated views on individual events. Reichenbach's 'fictitious' probabilities are an obviously not very satisfactory attempt to have it both ways: individual events do have a probability but not a real one. Salmon's account, which tries to remedy this inconsistency, goes wrong in a subtler but similar way: for him 'probability is literally and exclusively a concept that applies to infinite sequences', and yet it is supposedly rules of application for that concept which cover the single case. The cause of this vacillation is quite simply the failure of these writers to appreciate the relational character of the probability of individual events.

In failing to see that the probability of the outcome of a particular experiment is a relation crucially involving a reference class which has long run frequencies, Reichenbach and Salmon are driven to find some other means of characterising the manner in which such outcomes have probabilities but only on the basis of frequencies in reference classes. A 'fictitious probability' which is a 'substitute constructed in terms of class probabilities' is Reichenbach's attempt at such a characterization; in Salmon this characterization is given in terms of a weight determined by rules of application for the concept of class probabilities. Unfortunately neither of these accounts provides an adequate, or even consistent, view of the actual nature of the relationship between the probability of a particular outcome and the class frequencies which determine that probability.

From this perspective it is easy enough now to remedy the inconsistencies we found in Reichenbach and Salmon. For Reichenbach this is a straightforward - we need only admit that his 'fictitious probabilities' derived from class frequencies are genuine relational probabilities involving those classes. For Salmon the matter is slightly more complicated for, with the benefit of Reichenbach's experience and Carnap's distinction between probability theory proper and rules of application thereof, he has constructed an account in which the inconsistency is more deeply buried.

On the frequency theory for the single case given here, probability is a relation between two individual events; as we have repeatedly seen, this leads to numerous different probabilities for any individual event. But, as we noted in connection with Carnap's theory, what we ordinarily require for the

purposes of action is a unique probability in a given situation. And, as in Carnap, it is just here that methodological considerations come into play; we use certain methodological principles to give unique application of our relational definition of probability.

Now as we will see in the next two chapters when we consider in detail the principle of choosing the narrowest available reference class, Salmon (building on ideas of Reichenbach) articulates the correct methodological principle for determining unique probabilities for individual events. This principle is then indeed a rule of application for the concept of probability, but only if probability is defined as a relational property of individual events, rather than a relation between classes of them. In short, the theory of probability which Salmon correctly gives rules of application for is not the standard frequency theory he explicitly advocates, but rather is the relational theory put forward in this chapter for the probability of individual events.

Although the treatment suggested by Reichenbach and Salmon of the single case can only be consistently worked out if the standard frequency theory as such is abandoned in favour of the one advocated here, it may be doubted if this latter theory can provide an adequate account of the class probabilities of which the standard frequency theory provided a clear and consistent account. For example, when we spoke of the probability of 'heads' with a particular coin - as opposed to heads on a given toss - it seemed quite plausible to accept the standard frequency theory's identification of probability with the limit of relative frequency of heads in repeated tosses. Of course one



might accept the standard frequency account as valid for some range of cases - class probabilities - and the alternative offered here as valid for other examples - single case probabilities. But here we are left with two distinct frequency concepts of probability and, of course, this is to be avoided if possible.

In point of fact our frequency theory for individual events is easily - and literally - generalized to cover the case of so-called class probabilities, and it is with a discussion of this topic that I will conclude the present chapter. Given that we have defined a probability function whose arguments are individual events, examples in which probability is apparently assigned to classes of individual events can easily be understood as the assignment of that probability to each member of the class in question. That is to say statements about class probabilities are to be understood as universal quantifications over probabilities ascribed to individual events.

The treatment of class probabilities as universal quantifications over individual events - which I will elaborate on shortly - can be traced back in one form as far as Von Kries, whose opinion to this effect is cited and rejected by Von Mises.<sup>1</sup> Ironically, however, Reichenbach's account of class probabilities is given in just this fashion<sup>2</sup>. Reichenbach begins by characterizing probability as a relation of implication between individual events, symbolized ' $\dots \Rightarrow \dots$ ' and read 'If  $\dots$  then  $\dots$  is probable to degree  $p$ '. Individual events (which are specified in the dots) consist for him of a member of a sequence of experiments  $x_1 \dots x_i \dots$  belonging to a reference class and a member of a sequence of experiments  $y_1 \dots y_i \dots$

belonging to an attribute class B. The sequence  $y_1 \dots y_1 \dots$  is possibly, but not necessarily, the same as the sequence  $x_1 \dots x_1 \dots$  (When the two sequences are the same, Reichenbach's symbolism collapses to the simpler one used here).

Reichenbach tentatively suggests

$$x_1 \in A \quad \Rightarrow_p \quad y_1 \in B \quad (2)$$

as the form of a probability statement. This corresponds exactly (save for the use of two variables  $x$  and  $y$ ) to our theory in which probability is assigned to an individual event  $x_n \in B$  in relation to another event  $x_n \in A$ . But having formulated (2), Reichenbach hastily adds:

However, even (2) does not completely represent the form of the probability statement; we must add the assertion that the same probability implication holds for each pair  $x_i, y_i$ . This generalization is expressible by two all-operators meaning, "for all  $x_i$  and for all  $y_i$ ". Using an abbreviation, we can reduce the two all-operators to one by placing only the subscript  $i$  in the parentheses of the operator. Thus the probability statement is written:

$$(1) \quad (x_i \in A \Rightarrow_p y_i \in B) \quad (3)$$

This expression represents the final form of the probability statement. The probability statement is a general implication between statements concerning a class membership of the elements of a certain given sequence.

This latter qualification concerning the addition of the universal quantifier is, of course, what leads him to endorse the standard frequency theory:

Therefore, strictly speaking, the probability implication must be regarded as a three term relation between two classes and a sequence pair. The pair of sequences provides the domain with respect to which the probability implication assumes a determinate degree.

However, and this is the important point here, the probability statement (3) is in fact most naturally understood as a form of universal quantification ascribing a probability  $p$  for an outcome of a certain kind ( $y_i \in B$ ) to each individual event  $x_i \in A$ . Thus the statements about class probabilities which apparently provide a certain justification for the standard frequency theory are in no way incompatible with the frequency theory for individual events put forward here. In our theory probabilities are assigned to individual events in pairs on the basis of class frequencies and, as Reichenbach's formulation clearly brings out, class probabilities can readily be treated as probabilities attaching to all pairs of certain individual events, i.e. the pairs  $\langle x_i \in A, y_i \in B \rangle$  for all  $i$  of a sequence of experiments. On our theory the numerical value assigned to each such pair would be the limit of the relative frequency of  $B$  on  $A$  and this of course accords with the numerical values indicated as appropriate for class probabilities by advocates of the standard frequency theory.

Thus, for example, when we speak of the probability of mortality in the next year among 40 year old men and say that this equals the percentage  $p$  of death over a year among this class, we are - on my view - making a statement about the probability of mortality for each member of that class of

men. We are saying that every man, in so far as he is 40 years old, has a certain chance,  $p$ , of death. Of course, any particular man of 40 might have other properties in relation to which his chance of mortality would vary, but ignoring such possibilities, and so relative simply to the fact that he is 40, each man has the probability  $p$  of dying within the next year. On this account of the matter, when we make a statement about the probability of mortality in a given class of men we are not ascribing to that class of men a property that no individual man can possibly have; rather (more plausibly) we are ascribing a property that any individual man can, and indeed will, have in so far as he belongs to the class in question.

Such an account of class probabilities assigns two roles to reference class of the standard frequency theory. First of all each individual experiment  $x_i$  of the reference class  $A$  is assigned a probability for an outcome (in our symbolism  $x_i \in B$ ). The probability for the individual events  $x_i \in B$  is not absolute but is relative to the events  $x_i \in A$ . This relational probability for all the pairs of individual events is nevertheless genuine and this, of course, is the difference between our theory and the standard theory. But secondly, the probability for the individual events  $x_i \in B$ , relative to  $x_i \in A$ , is to be measured by the limit of  $B$  on the reference class  $A$  and this, of course, is a point of similarity between our theory and the standard one. To put it generally, a class probability is, for me, a probability for an outcome for every individual experiment of a class, relative to the fact the experiment belongs to that class. The numerical value of probability for this outcome is



determined by the frequency of that kind of outcome in the reference class.

For the standard theory class probabilities are just these frequencies.

This treatment of class probabilities in our theory has been dwelt on at length for two reasons. First of all it *prima facie* provides an additional argument in favour of our frequency theory for individual events, for our theory is shown capable of giving a satisfactory account of the class probabilities that figure so prominently in the standard theory. But of course the converse does not hold, for, as we have seen, the standard frequency theory cannot provide an adequate account of probabilities for individual events. Thus on the grounds of generality alone our theory for individual events is to be preferred.

Secondly I think this argument can be extended to explain why many, following Von Mises, have further held that we can only legitimately speak of a class probability if the class in question fulfills the condition of randomness. Exploration of this point must, however, await further discussion in the next chapter of a procedure for determining unique probabilities for individual events on the basis of the admittedly relational definition of probability given in this chapter.

**Footnotes    Chapter II.**

1.     Von Mises, R., *op. cit.* p.18.
2.     Reichenbach, *op. cit.*, p.46 ff.

### CHAPTER III

#### Randomness and Unique Probabilities

While the definition of probability offered in the previous chapter is genuinely applicable to individual events, it is abundantly clear that this definition does not provide a criterion for assigning unique probability values to individual events. In the above definition the widely recognised variability of the probability of individual events on the reference class chosen is not avoided, rather that variability is simply taken to be a result of the relational character of probability. Accordingly in so far as it is accepted as an adequacy condition on definitions of probability that they warrant the assignment of unique probability values, the definition of probability given in the previous chapter will be found inadequate. In this chapter I wish to investigate both the details and consequences of modifying the above definition to assign unique probability values to individual events.

I will begin this chapter with a preliminary version of such a modification and then investigate the relationship between it and Von Mises' concept of randomness. This study of this relationship is of some independent interest and will be seen to support persuasively the legitimacy of a frequency concept of probabilities for individual events. In addition the study of this relationship will provide us with a perspective from which we may constructively

approach in the next chapter various difficulties that will have arisen for our preliminary version of a principle for assigning unique probabilities to individual events.

As we have seen with quotations from Reichenbach, Popper and Salmon, the most natural means for those of a frequentist persuasion to determine unique probabilities for individual events is to rely on the frequencies found in the narrowest reference class for which statistics are available. As a historical point we should note that this procedure was sufficiently appealing to have attracted Venn's attention to one form of it in the 19th Century and Keynes, following Venn, considers it at some length in his Treatise. In the next two chapters I will refer to this procedure as that of selecting the narrowest available reference class for this is a convenient manner of expression and serves to stress the parallels to the procedure of using the total available evidence.

If we wish to have a definition of probability which assigns unique probabilities to individual events, we simply amend the definition offered in Chapter II in light of the procedure of choosing the narrowest available reference class. In Chapter II any infinite class to which an experiment belonged could serve as the reference class required by our theory of probability; in order to overcome the variability of probability values so determined we would reformulate the definition by limiting the possible reference classes to those that constitute the narrowest available reference class.



In point of fact it is a matter of considerable controversy<sup>2</sup> whether an adequate definition of probability must determine unique probabilities - if probability is accepted as a relational concept there is no inconsistency in offering a definition that yields variable or relational probabilities. However, even if we think an adequate theory of probability may assign variable or relational probabilities, it must be admitted that the practical need for a unique response in the face of uncertainty presses us to supplement any relational theory of probability with an explanation of how it can be applied in particular situations to yield unique results. This, as we saw in the previous two chapters, is one role envisioned by Carnap for methodological rules of application and I shall follow him here in this. Thus rather than abandoning the relational definition of probability offered in Chapter II, I shall continue to use that definition and treat the principle of choosing the narrowest available reference class as a methodological rule, required by our practical needs, for determining unique probabilities for individual events. (This, of course, was just the status intended for an equivalent principle advocated by Salmon; unfortunately, as we saw, the definition of probability expounded by Salmon did not apply to individual events and so, strictly speaking, his rule of application could not yield the desired results).

As a start we may formulate the methodological principle of choosing the narrowest available reference class thus in terms of the definition offered in Chapter II:

In ascribing a probability to the individual event  $x_n \in B$  in relation to the event  $x_n \in A$  choose only individual events  $x_n \in A$  for which the reference class  $A$  fulfills the following conditions:

- (1) The class  $A$  must be (potentially) infinite and the existing observational data sufficiently extensive to justify extrapolating current observations of relative frequency of  $B$  on  $A$  to a limit.
- (2) The defining characteristic of the class  $A$  must be of such a kind that we can ascertain of the individual experiment  $x_n$  whether or not it fulfills that characteristic prior to knowledge of the occurrence or non-occurrence of  $x_n \in B$ .
- (3) The class  $A$  must not contain any proper sub-class fulfilling conditions (1) and (2), except where the limit of the relative frequency of  $B$  on that sub-class = the limit of  $B$  on  $A$  itself.

Save for a problem to be raised as (d) at the start of the next chapter, conditions (1) - (3) are intended as a preliminary but plausible formulation of the necessary and sufficient conditions for a reference class being the narrowest available. Condition (3) contains the definition of narrowness as well as the usual stipulation that a narrower class with unchanged limits be disregarded.

Condition (2) is intended to rule out such 'unavailable' narrower reference classes as, e.g., the class of tosses of this coin that lands heads up. Clearly unless we can ascertain of an experiment  $x_n$  that it fulfills the defining characteristic of  $A$  prior to knowledge of the outcome of  $x_n$ , we can not 'avail' ourselves of  $A$  for determining the probability of the outcome of  $x_n$ . As it

stands condition (2) allows a large number of 'available' reference classes for it only requires that the defining characteristic of the class allow us to be able to ascertain that  $x_n$  fulfills it i.e. belongs to A, prior to knowledge of the occurrence or non-occurrence of  $x_n \in B$ . On the one hand, condition (2) as it stands insures the objectivity of our methodological principle, for the failure of a particular person actually to ascertain that  $x_n \in A$  prior to the occurrence or non-occurrence of  $x_n \in B$  does not entail that A was unavailable for him; condition (2) permits as available reference classes those classes to which  $x_n$  belongs and for which it is possible to determine, prior to knowledge of the eventual result, that  $x_n$  does indeed belong to that class.

On the other hand, condition (2) gives rise to two difficulties.

First of all it may be doubted whether all reference classes that fulfill condition (1) can be deemed available simply on the basis of also fulfilling condition (2) as it stands. For example, consider the case of a horse that has a rare and as yet undetected disease. This disease is known to affect running performance to a determinate degree (thus allowing condition (1) to be fulfilled) and, moreover, can be diagnosed quite early if the proper tests are administered. In using the principle of choosing the narrowest available reference class to determine the correct probability of this horse winning a particular race, are we to use as the relevant statistic the roughly calculated frequency of victory in the class of horses of similar blood line, previous form etc? Or rather are we to regard the narrowest available reference class as the one containing all horses of similar blood line, current form etc., who also have this rare disease.

Condition (2) as formulated above indicates the latter as correct, for the disease, while rare and undetected, could in theory be diagnosed without knowing the result of the race if the correct tests were made. This conclusion seems not wholly implausible but is debatable, as we could hardly expect even a well informed punter to be aware of what is the narrowest available reference class here; however, for the purposes of convenience, debate over this point will be postponed until the next chapter. As a preliminary formulation condition (2) will suffice and while the results of applying it may set standards for determining probabilities correctly that appear high for punters or bookmakers, these standards are not demonstrably inappropriate for the more exact sciences.

Similarly we may note another difficulty with condition (2) which, however, will also only be discussed at length in the next chapter. The objectivity insured by condition (2) is purchased at the expense of rigour, for as it stands condition (2) relies on the rather vague concept of a defining characteristic of 'such a kind that we can ascertain of any individual experiment  $x_n$  whether or not it fulfills that defining characteristic prior to knowledge of the occurrence or the non-occurrence of  $x_n \in B$ '. Although the intention of this criterion is clear, it is far from obvious if it has a wholly determinate application. The obscurity implicit in the vague formulation of condition (2), or any similar criterion, presumably accounts for some of the misgivings often expressed concerning the principle of choosing the narrowest available reference class. We shall return to this problem in the next chapter. However, for the moment we are not concerned with resolving all difficulties with the principle of choosing the narrowest available reference class; rather we are only spelling out in a



preliminary fashion a plausible formulation of what is generally, if a bit vaguely, understood by that principle - and for this purpose condition (2) as it stands seems adequate.

Condition (1) obviously formulates part of what is usually understood by an available reference class; in particular, for an infinite class with a limit of relative frequency to be 'available' to us for measuring probabilities by that limit we must be able to estimate the limit from our always finite data. While less vague than condition (2), condition (1) obviously requires further comment. However, insofar as it does not inhibit eventual discussion of the problem of applying idealized conceptions to empirically given situations, it seems both desirable and expedient in constructing a frequency theory of probability to develop as much of it as possible in isolation from the problem of extrapolation to infinity from finite observational data. Accordingly, the remainder of this chapter will only deal with ideal sequences with known limits; thus we may assume that we begin with a specific set of infinite sequences with known limits both for the main sequence and any sub-sequence we are concerned with. Condition (1) is then replaced by (1'); (1') The reference class A must be an infinite sequence with known limiting properties.\*

---

\*The phrase 'with known limiting properties' is not meant to imply that we are restricted to infinite sequences in which a genuine limit is approached. It is possible that our narrowest available reference class will be one which we know has an oscillating value of relative frequency. For such a case we must stipulate that the individual event in question can not be assigned a unique probability.

Turning our attention to the concept of randomness, we may recall that Von Mises defined a reference class (collective) of experiments as random in respect to an attribute class  $B$  if and only if there existed no mathematically defined infinite sub-sequence of the collective in which the limit of  $B$  differed from the limit of  $B$  in the collective itself. The mathematical rule which determined the selection or deletion of experiments  $x_i$  of the collective in order to form the infinite sub-sequence, must be such that the selection or deletion of  $x_i$  is independent of the result  $B$  or  $\sim B$  for  $x_i$ , that is, depends only on the number  $i$  and the results of the previous experiments  $x_1 \dots x_{i-1}$ .<sup>3</sup> To this definition of randomness Von Mises added the restriction that only random sequences were probability sequences - his so-called requirement of randomness.

Reichenbach, for one, disputed Von Mises' requirement of randomness<sup>4</sup> and in Von Mises' original presentation of it as an additional requirement on the standard frequency theory, it is difficult to see anything more than a vague, albeit strong, intuitive plausibility for the requirement. Von Mises himself claimed that his requirement of randomness - also known as the Principle of the Impossibility of a Gambling System - was obtained by generalizing on the experience of games of chance as, for example, those carried on by gambling banks and insurance companies. This justification for his requirement of randomness, of course, accords with Von Mises' view of probability theory as an empirical science, just like, say, mechanics. Whatever the merits of this positivist account of probability theory - and like Carnap and Waismann<sup>5</sup>, I for one, would dispute it - such a view can not be used to justify his requirement of randomness.

In the first instance the claim that only random sequences are probabilistic is circular if based on a generalization of games of 'chance', for 'chance' phenomena are not specified independently of our intuitive identification of chance and probability. Admittedly Von Mises often states that he has simply delineated a set of empirical phenomena, the study of which he proposes to call probability theory.<sup>6</sup> One characteristic by which this set is defined is the property of randomness and, of course, one can not deny anyone the right to specify a set of phenomena and then set about studying the properties and relations of those phenomena. But one may deny that such a study corresponds to what has traditionally been known as probability theory or that such usage of the term 'probable' coincides with our ordinary usage. As Reichenbach pointed out, many sequences ordinarily called probabilistic do not obey Von Mises' requirement, e.g., the sequences of automobile accidents in cities for which the frequency will vary regularly with days of the week.

Von Mises' reply to this objection was that his theory was truly scientific and so, of necessity, had to diverge at certain points from ordinary usage and previous unscientific studies of related phenomena. However, it is widely known that the frequency theory can be given a perfectly exact formulation without the requirement of randomness and, moreover, that all the fundamental theorems (e.g. the Laws of Large Numbers) can be proved without the requirement of randomness. In addition, as Reichenbach again pointed out, there are many examples of non-random sequences in physics

which are known as probability sequences and it is obviously desirable for these to be treated in any 'scientific' theory of probability.

However, and this is the central point of this chapter, we may give a perfectly clear justification for Von Mises' requirement of randomness if we abandon the standard frequency theory to which it was originally added and instead accept the frequency theory for individual events proposed in Chapter II, with the additional restrictions on reference classes embodied in conditions (1') - (3) to assign unique probability values. The justification for the requirement is now as follows: we have found that probability of the individual event  $x_n \in B$  will vary as the experiment  $x_n$  is thought to belong to different reference classes. To overcome this we appeal to the principle of choosing the narrowest available reference class i.e. a class fulfilling conditions (1') - (3).

Now let us assume that a particular reference class  $A$  with a limit  $p$  for  $B$  presents itself to us as the narrowest available one to measure the probability of  $x_n \in B$ . Now, and this is crucial, if  $A$  is not random with respect to  $B$ , it is quite possible that, contrary to our assumption,  $A$  is not the narrowest one available for measuring the probability of  $x_n \in B$ . If  $A$  is not random with respect to  $B$ , there will be some rule that will select out an infinite sub-sequence  $A_1$  in which  $B$  approaches a limiting value of relative frequency not equal to  $p$ . Assuming  $x_n$  belongs to that infinite sub-sequence  $A_1$ , then  $A_1$  will constitute for the event  $x_n \in B$  an available reference class narrower than  $A$ . The sub-sequence with a different limit is, by definition, infinite and a proper sub-class of the initial reference class  $A$ . All that is further required



for this infinite sub-sequence  $A_1$  to which  $x_n$  belongs to constitute a narrower available reference class for measuring the probability of  $x_n \in B$  is that we be able to ascertain that  $x_n$  fulfills the defining characteristic of  $A_1$  prior to knowledge of the occurrence or non-occurrence of  $x_n \in B$ . But this is guaranteed by the definition of randomness, for randomness is defined in terms of selections and deletions of  $x_i$  by rules based only on the number  $i$  and the results  $x_1 \dots x_{i-1}$ . To put it more explicitly, the defining characteristic of an infinite sub-sequence  $A_1$  that violates the randomness of the initial sequence  $A$  will be a mathematical rule for selecting elements  $x_i$  based only on the index number  $i$  and the results  $x_1 \dots x_{i-1}$ . Thus the question of whether or not a particular experiment  $x_n$  belongs to  $A_1$  can be settled on the basis of only the number  $n$  and the results  $x_1 \dots x_{n-1}$  i.e. prior to knowledge of the occurrence or non-occurrence of  $x_n \in B$ . Therefore, if  $x_n$  is a member of an infinite sub-sequence of the kind that makes initial sequence  $A$  non-random, that sub-sequence will constitute a reference class narrower than  $A$  and one which is available for measuring the probability of  $x_n \in B$ .

Of course as not all  $x_i$  of  $A$  need belong to the infinite sub-sequence (or sub-sequences if there are more than one) which violate the randomness of  $A$ , the above analysis will not necessarily apply to the particular  $x_n$  of  $A$  we are at any moment concerned with. However it is clear that some, indeed an infinity, of  $x_i$  of  $A$  will belong to the sub-sequence (or sub-sequences) which violate the randomness of  $A$ . Accordingly any non-random reference class  $A$  can not be the narrowest reference class available for all its individual elements.

Thus the requirement of randomness insures that we only use those sequences that are suitable for determining unique probabilities for all of their constituent experiments.

These considerations go a long way towards accounting for the interest and significance of Von Mises' concept of randomness as well as his emphatic insistence that only random sequences are truly probabilistic. As originally presented in terms of the standard frequency theory, Von Mises' requirement of randomness had at best a certain intuitive appeal; re-interpreted as above in terms of a frequency theory designed to give unique probabilities to individual events, its rationale becomes perfectly clear. Despite Von Mises' and others' explicit claims, ordered pairs consisting of an infinite reference class and attribute class are not what we ordinarily deem probable, rather it is ordered pairs of individual events that we hold probable on the basis of frequencies in infinite sequences of experiments. And, in light of the most natural criterion for determining unique probabilities, only frequencies in random sequences of experiments can suffice for determining unique probabilities for all the experiments of the sequence. Random sequences are required insofar as we are concerned with assigning unique probabilities to every member of a sequence on the basis of a frequency in that sequence; thus in constructing a theory of probability which assigns unique probabilities to the individual members of a sequence we will, like Von Mises, regard only random sequences as truly probabilistic for only they can be used to assign unique probabilities to all their members.

Let me express this point in another way: in Chapter II we saw that statements about class probabilities were to be construed as universally quantified statements assigning a (relational) probability  $p$  for the occurrence of an attribute  $B$  to each experiment  $x_i$  of a given class  $A$ . The frequency of  $B$  in this class also served to measure the probability of  $B$  for each experiment  $x_i$ , relative to the fact that  $x_i$  was a member of  $A$ . Now it can easily be shown that Von Mises' claim that 'class probabilities' only obtain when we have random classes follows directly from this analysis of class probabilities, given only the further proviso that we wish to assign unique probabilities by using the narrowest available reference class. Class probabilities, on the analysis offered, are probabilities for each and every member of a class numerically equal to the frequency found in that class. And, as we have just seen, it is only frequencies in random classes that can be assigned as a unique probability to each and every constituent member of a class - an infinite sub-sequence that violates the randomness of an original collective constitutes a reference class with a different limit, narrower than the original collective, and one that, moreover, is 'available' for determining the probability of some of the experiments of the original collective.

Thus Von Mises' celebrated requirement of randomness can be seen as a direct consequence of the frequency theory for the single case advocated in Chapter II, when that theory is supplemented by a methodological principle for determining unique probabilities. Given that little, if any, other argument can be given to justify the requirement of randomness (we saw how

indecisive Von Mises' original arguments were), it seems plausible to think that the widespread acceptance of it depends on this (admittedly unrecognized) status as a consequence of a frequency theory designed to assign unique probabilities to individual events.

This conclusion extends to the standard frequency theory with the requirement of randomness the line of argument already put forward for the theory without the requirement - that our theory is the more general in so far as examples which figure prominently in the standard theory can be understood as special instances of our theory. In Chapter II we saw how class probabilities could be understood as the special case of relational probabilities for an outcome assigned to each of a class of individual experiments. Now we see that Von Mises' proposal to assign class probabilities only to random classes can be understood as deriving from the special case in which we desire to assign unique probabilities to each of a class of individual experiments.

Curiously enough this line of argument concerning the standard frequency theory with the requirement of randomness is, to some extent, supported by various passages from Reichenbach and Salmon. They, like other writers<sup>7</sup>, perceived that there is a connection between randomness and ordinary statements in which probabilities are apparently ascribed to the single case. In discussing the question of the 'weight' to be given to scientific hypotheses, Reichenbach returns to the question of choosing a suitable reference class. He correctly sees that such hypotheses are simply one instance of the problem of the single case. Reichenbach suggests that for scientific hypotheses



we should choose a 'homogeneous' reference class. This is explained thus:

A class of tuberculosis is a homogeneous reference class. But it would seem unwise to compile death statistics in a class of persons with different diseases or in a class including human beings and animals. The definition of the predicate 'homogeneous' depends on the state of our knowledge. An inhomogeneous class can be defined as a class for which we know methods by means of which the class can be so subdivided without the use of the attribute (see Section 30), that subclasses of different frequencies for the attribute result. The subdivision will sometimes be achieved by reference to other attributes such as are given in the example of different diseases or biological species. However, it can be achieved also by dividing the total sequence, or ordered class, into consecutive sections.<sup>8</sup>

The section 30 referred to is devoted to a discussion of Von Mises' requirement of randomness and although Reichenbach nowhere explicitly discusses the relationship between randomness and homogeneity, it is clear that he held them to be connected concepts. One difference is apparent - the procedures of subdivision that define homogeneity include selection by physical attributes, whereas randomness, of course, is defined solely in terms of mathematical rules of selection. Moreover as homogeneity is defined in terms of known methods of subdivision, it is unclear whether all the mathematical rules of selection by which randomness is defined would figure in Reichenbach's definition of homogeneity.

Salmon, as we have seen, developed Reichenbach's suggestion here systematically and in fact proposed in general to replace the principle of selecting the narrowest available reference class with the equivalent formulation

of selecting the broadest homogeneous reference class. Again just referring to Von Mises' concept of randomness, without any elaboration, Salmon proceeds to define homogeneity exactly as Reichenbach did - save that for him homogeneity is not relativised to our knowledge, i.e. a class may be inhomogeneous and yet, as he puts it, we may 'not know of a way to make a relevant partition of it'.<sup>9</sup> This additional proviso makes it fairly clear (as it was not in Reichenbach) that, for Salmon, all homogeneous classes will be random in Von Mises' sense, since if I read him correctly, all mathematical procedures of selection envisioned by Von Mises are included in his definition of homogeneity. Salmon further states that in assigning a weight to the single case we are to employ the broadest homogeneous reference class for this ensures that we consider the maximum number of relevant cases in our estimation of frequencies ('broadest class' here means that including the most relevant elements).

Now Reichenbach's proposal to use homogeneous reference for determining the 'weight' of scientific hypotheses and Salmon's proposal to use the broadest homogeneous reference class for determining weights for all single cases both capture something of the relationship between randomness and single case probabilities I have elaborated here and so their proposals support, in some rough degree, the line of approach taken in this chapter. But there is a crucial difference between what Salmon and Reichenbach, on the one hand, and I, on the other, perceive to be the relationship between randomness and the single case and it is with a comparison of our positions here I will close this chapter.

On their accounts we are not given - and could not be given - any explanation of why Von Mises, or anyone else, might think that only random classes are truly probabilistic - the so-called requirement of randomness. On their account, probability indeed only applies legitimately to classes but no additional requirement of randomness is imposed. Randomness then figures as part of a concept, (homogeneity) used in the verbal (or rather written) expression of a rule for determining 'weights' for single cases.

But on the view advocated here, probability is legitimately (and solely) a relational property of individual events and the concept of a class probability is to be construed in terms of universal quantification over these probabilities. If we desire unique probabilities for individual events we are required to adopt a principle for selecting a unique reference class for an individual event; for this purpose the principle of choosing the narrowest available reference class is both long standing (pre-dating the works of Von Mises, Reichenbach and Salmon) and highly plausible. It then can be shown that this principle conjoined to our frequency theory for individual events and its explanation of so-called class probabilities entails that only random classes display class probabilities. This, furthermore, suggested that it was a tacit acceptance of such a theory for the single case which all along made Von Mises' requirement of randomness seem so plausible.

So the account given by Salmon and Reichenbach of weights (rather than probabilities) for the single case, determined by homogeneous

reference classes, misses the essential point: by taking probability to apply only to single cases (or in quantified statements classes of them) we may give a full justification of the widely accepted requirement of randomness. To repeat, as advocates of the standard frequency theory, Salmon and Reichenbach missed (and had to miss) the essential point that Von Mises' requirement of randomness is a direct consequence of a frequency theory designed to assign genuine and unique probabilities to the single case.



## Footnotes Chapter III

1. Keynes, J.M., *op. cit.*, pp. 103 ff.
2. Carnap, R., The Logical Foundations of Probability, 2nd ed., Chicago 1962, Section 44, Ayer, A.J., Probability and Evidence London 1972 pp. 57-58, Mellor, D.H., The Matter of Chance, Cambridge 1973, p. 60, and Watling, J., *Chance*, *Arist. Soc. Suppl.* vol. 73, pp. 37-48, among many others.
3. Von Mises, R., *op. cit.*, pp. 23-28 and 50; see also Church, A., *On The Concept of a Random Sequence*, *Bull. Am. Math. Soc.* vol. 46, in which this specific formulation of Von Mises' suggestion is put forward. Recent work suggests that this formulation may be somewhat narrower than Von Mises' original wording permits (Review of Lovelund, A., *JSL*, vol. 38, p. 357), but as this formulation has long been accepted it will be used here.
4. Reichenbach, H., *op. cit.*, Sections 27 and 30.
5. Carnap, R., *op. cit.*, p. 34.
6. Von Mises, R., *op. cit.*, p. 26.
7. Mackie, J.L., Truth Probability and Paradox, Oxford 1973, p. 177 and Kyburg, H.E., *Probability and Randomness*, *Theoria*, vol. 29, 1963, pp. 27-55.
8. Reichenbach, H., *op. cit.*, p. 443.
9. Salmon, W.C., *op. cit.*, p. 92.

## CHAPTER IV

### Choosing The Narrowest Available Reference Class

As the line of argument in Chapter III depends heavily on the principle of choosing the narrowest available reference class, it is essential that we subject this principle to careful critical scrutiny. In Chapter III we noticed three *prima facie* difficulties with the provisional formulation of it in clauses (1'), or (1), - (3). Let me review these and mention, for good measure, two other problems. The problems already noticed are:

- (a) Clause (1'), as opposed to (1), relies on the unrealistic and idealized conception of infinite sequences whose limiting properties are taken as known. Although more realistic, clause (1) relies on an unexplained concept of extrapolability which requires further explanation and elucidation.
- (b) Clause (2) sets particularly high standards for the range of classes deemed available for determining probability values. As our example of the diseased horse showed, there is little reason to expect even a prudent investigator to be able to identify invariably the narrowest reference class available in the sense of (2).
- (c) The standards set by (2) are not only extremely high but also quite vague, relying on an unexplicated concept of knowledge that 'could be ascertained' prior to knowing the outcome of an experiment, even if it is not.

In addition to these difficulties we should note that, contrary to the impression given in Chapter III, there is no guarantee that clauses (1'), or (1), - (3) are sufficient to determine just one class to be the narrowest one available for a given individual experiment. Consider a somewhat unusual sequence of tosses of a coin - one for which we know the limiting value of heads in prime numbered tosses to be, roughly, 90%. The limit on even tosses is known to be, roughly, 40%. What probability should be assigned to heads on the second toss, which is both even and prime? The difficulty here is that both the class of prime tosses and that of even tosses fulfill clauses (1') - (3) for the second toss, but each have different limits for heads.

The existence of a number of different reference classes  $A_1 \dots A_k$  all fulfilling conditions (1'), or (1), - (3) for a particular outcome  $x_n \in B$  would be unproblematic if each such reference class had the same limit of relative frequency for the attribute B, since we could simply ascribe this limit  $p$  as the probability of the event  $x_n \in B$ . However, if there are different limits of relative frequency of B on the classes  $A_1 \dots A_k$  fulfilling conditions (1'), or (1), - (3) we can not assign a unique probability value to  $x_n \in B$ .

While there seems nothing problematic about formulating a methodological principle intended to determine unique probabilities that fails to do so in a certain minority of actual cases, clauses (1'), or (1), - (3) obviously should be re-formulated in some way so as to minimize this possibility. We may call this problem

- (d) That of finding a re-formulation of clauses (1'), or (1), - (3) which maximizes the chance of determining a unique probability.

Finally we should note

- (e) Ayer's objection. Ayer has argued that, although the policy of choosing the narrowest available reference class is correct, there can be no justification for it within a frequency theory for the single case<sup>1</sup>. Ayer, without spelling out in detail any frequency theory for the single case, clearly has in mind a theory similar to that advocated in Chapter II; his objection is that such a theory can not justify the methodological principle in question.

We may begin with (e) and work our way back to (a). It is not altogether clear why Ayer requires a frequency theory for the single case to justify a rule for determining unique probabilities. Some have disputed this requirement in general<sup>2</sup>; in any case, on the construction given here, the policy of choosing the narrowest available class is a principle of application prompted by practical needs. All that is required to justify this policy is a demonstration that adopting it leads to desirable results. Here Salmon, again following Reichenbach, has quite correctly pointed out that the policy of determining betting odds on the basis of the frequency in the narrowest available reference class is highly desirable, for anyone who adopts it will gain in the long run at the expense of those who offer higher odds or accept lower ones.

If we know the long run probabilities, a certain type of success is assured by following the methodological rules for handling single events. A given individual deals with a great variety of single events.... It is demonstrable that he will be successful 'in the long run' if he follows the



methodological rules laid down. He will win a few and lose a few, but if he so acts that he has a positive expectation of gain in each case, he will come out ahead in the long run.<sup>3</sup>

The basic point here is readily illustrated by a favourite example of frequentists, that of an insurance company attempting to maximize its profits. With Salmon we may assume that the company has extensive knowledge of, say, long run mortality rates for numerous categories of potential policy holders. It knows, for example, that the long run rate of mortality among the class  $A_1$  of 40 year old men passing a particular medical test is, say,  $p_1$ . In addition it knows that the mortality rate among the sub-class  $A_2$  of  $A_1$  of 40 year old men who smoke heavily is  $p_2$ , where  $p_1 < p_2$ . Similarly it knows  $p_3$  to be the rate among the sub-class  $A_3$  of 40 year old smokers who also work in a particularly stressful occupation;  $p_3$  here is greater than  $p_2$ . The company now must decide on a premium for life insurance to charge members of the class  $A_3$ . To fix a particular premium is, of course, to offer certain betting odds on a man's life, although this is a rather callous way of putting it.

It should be obvious that the company will lose in the long run if it adopts the lower mortality rates  $p_1$  or  $p_2$  instead of  $p_3$  to fix premiums for members of the class  $A_3$ . Similarly it would gain in the long run if it were able to find clients among members of  $A_3$  who would pay higher premiums based on a mortality rate higher than  $p_3$ . Thus as remarked above, fixing betting odds, or premiums, on the basis of the long run frequency in the narrowest available reference class - here  $A_3$  for the members of  $A_3$  -

leads to long run profit at the expense of those who do not so use the narrowest available reference class. More generally, it is the case that transition to the frequency found in a narrower reference class will increase one's long run gain, or decrease one's loss, even if the frequency in the narrowest available reference class is for some reason not used - an insurance company using  $p_2$  to determine the premium for members of  $A_3$  would lose less than one using  $p_1$ .

One point should be made about this last remark that transition to frequencies in a narrower class will increase one's long run gain even if the narrowest available class is not itself used. Ayer might reply by pointing out that this hinges on the assumption that the frequencies found in the progressively narrower reference classes become progressively closer and closer to each other. That is, in our example  $p_1 < p_2 < p_3$  and so, in particular  $p_2$  is closer to  $p_3$  than is  $p_1$ . If, instead, we had, say,  $p_1 \neq p_2$  but  $p_3 = p_1$ , there would be no gain in using the frequency  $p_3$  as found in  $A_3$  for members of  $A_3$  rather than  $p_1$  as found in  $A_1$ . Moreover, there would actually be a long-run loss involved if one used the frequency found in  $A_2$  rather than that in  $A_1$  for determining premiums for members of  $A_3$ . In general, unless the frequencies found in progressively narrower reference classes become closer and closer, use of progressively narrower reference classes can not be relied on to maximize long-run gain.

The answer to such a rejoinder on behalf of Ayer can be derived from Reichenbach's original discussion of the principle by choosing the narrowest available reference class, part of which was cited in Chapter I. In a passage not already quoted Reichenbach points out:

According to general experience the probability [i.e. relative frequency of the attribute] will approach a limit when the single case is enclosed in narrower and narrower classes, to the effect that from a certain step on further narrowing will not result in noticeable improvement.

The approach to a limit - or rather apparent approach to a limit as this is all we can observe - for frequencies found when transition is made to progressively narrower reference classes of course insures that the frequencies in such classes eventually become closer and closer to each other. Indeed 'approach to a limit' is an exact way of describing a progressive narrowing of differences. Thus the kind of change in frequencies required to justify transition to narrower and narrower reference classes is actually found in experience.

Our next problem (d) is at first glance quite formidable. In fact it can easily be shown that clauses (1'), or (1), - (3) as they stand never determine a unique reference class for an individual event. Any individual experiment  $x_n$  belongs to every class  $\phi$  where  $\phi = \psi \cup \{x_n\}$ .  $\psi$  here can take any value and this is the heart of the difficulty. Consider, for example, the toss of a regular die. To determine the probability of it landing face 6 up on toss  $x_n$  we will estimate the frequency of 6's in the class A of repeated tosses of regular die. We may assume A is observed to have a frequency of  $\frac{1}{6}$  for 6 and fulfills clauses (1) - (3), thus (apparently) being the narrowest available reference class. But the toss  $x_n$  also belongs to the quite arbitrarily constructed class  $A' \cup \{x_n\}$ , where  $A'$  is the repeated toss of a die which has six faces marked 6. Clearly  $A' \cup \{x_n\}$  fulfills clauses (1) - (3) for it contains no infinite subsequence  $A''$  of

which  $x_n$  is member and in which we estimate the long run frequency of 6 is different from that in  $A'$ . We now have two reference classes  $A$  and  $A' \cup \{x_n\}$  for  $x_n$  and, crucially, neither is a sub-class of the other. Rather than a unique narrowest available reference class, we have two reference classes with different frequencies, each, as it were, equally narrow.

The problem here is general and, moreover, is just one of a more extensive family of difficulties - examples involving Goodmanesque predicates would raise similar issues. Hempel, in a closely related context,<sup>4</sup> has suggested that such difficulties can be overcome by restricting ourselves to use of reference classes based on 'lawlike' predicates. The reference class  $A' \cup \{x_n\}$  would be ruled out on this proposal, as would many other problematic reference classes.

Although the concept of a 'lawlike' predicate is notoriously difficult to define, this is a problem endemic to the philosophy of science as a whole and thus should not be seen as an insurmountable obstacle for our theory in particular. Accordingly, as an admitted over-simplification, we may incorporate into our clause (2) the additional requirement that the defining characteristic of any admissible reference class be expressed in terms of a lawlike predicate.

However, even if we restrict our attention to classes based on 'lawlike' predicates, it has been argued that conditions such as (1) - (3) will not usually yield unique assignments of probability. To establish this Kyburg<sup>5</sup>,



for one, produces the following example: we wish to determine the probability that an object  $n$  belonging to the class  $C_1 \cap C_2 \cap C_3$  has the property  $G$ . Our relevant statistical knowledge is

- (1) The proportion of  $C_1$  that is  $G$  lies in the interval (0.77, 0.82)
- (2) The proportion of  $C_2$  that is  $G$  lies in the interval (0.70, 0.80)
- (3) The proportion of  $C_3$  that is  $G$  lies in the interval (0.78, 0.81)
- (4) The proportion of  $(C_1 \cap C_2)$  that is  $G$  lies in the interval (0.70, 0.85)
- (5) The proportion of  $(C_1 \cap C_3)$  that is  $G$  lies in the interval (0.75, 0.95)
- (6) The proportion of  $(C_2 \cap C_3)$  that is  $G$  lies in the interval (0.75, 0.85)
- (7) The proportion of  $(C_1 \cap C_2 \cap C_3)$  that is  $G$  lies in the interval (0.00, 1.00)

The difficulty here is that we lack knowledge of the relative frequency of  $G$  in  $(C_1 \cap C_2 \cap C_3)$  which would be the narrowest possible reference class; moreover the value of frequency among the equally narrow classes  $C_1 \cap C_2$ ,  $C_1 \cap C_3$ ,  $C_2 \cap C_3$  varies considerably.

I remarked above that such cases would not present a serious problem for our methodological principle if they occurred rarely since our principle is, after all, only a general rule to facilitate practical decisions. Kyburg, however, contends that this kind of problem 'constantly confronts the scientist and man of affairs, as well as, in somewhat foggy form, the ordinary citizen.' But is this actually so?

In Kyburg's example we begin with classes  $C_1$ ,  $C_2$ ,  $C_3$ , all with frequencies for  $G$  ranging within fairly narrow bounds around points between .75 and .80 - this much is quite plausible, for we will indeed find statistics where frequencies for different classes range within narrow bounds around some common points. But when we consider the narrower classes  $C_1 \cap C_2$ ,  $C_1 \cap C_3$ ,  $C_2 \cap C_3$ , we find the intervals within which our frequencies lie becomes much wider, e.g. the proportion of  $G$  in  $C_1 \cap C_3$  lies in the interval

(0.75, 0.95) while the widest interval among our original classes was (0.70, 0.80).

At first glance this may appear plausible enough - to move from frequencies in  $C_1$  and  $C_3$  separately to the frequency in  $C_1 \cap C_3$  we must find cases that are both  $C_1$  and  $C_3$ . Perhaps these are a bit uncommon and so our observed results will be accepted with the qualification that the actual frequency in  $C_1 \cap C_3$  may only coincide roughly with our limited initial observation. But on more careful scrutiny something rather curious has gone on in the transition from statistics for  $C_1$  and  $C_3$  to the statistic for  $C_1 \cap C_3$ . We believe that the proportion of G's in  $C_1 \cap C_3$  may be as high as .95. As it is known that the proportion of G's in  $C_1$  and  $C_3$  separately is never higher than .82, this belief must be based on observation of an increase in the incidence of G among those members of  $C_1$  and  $C_3$  that are members of  $C_1 \cap C_3$ . On the other hand we are also supposed to believe that the proportion of G's in  $C_1 \cap C_3$  may be as low as 0.75. As it is known that the proportion of G's in  $C_1$  and  $C_3$  separately is not less than 0.77, this belief must be based on some indication of a small, but perceptible, decrease in the incidence of G's among those members of  $C_1$  and  $C_3$  which belong to  $C_1 \cap C_3$ . But these two assumptions about the change of incidence of G in transition from  $C_1$  and  $C_3$  separately to  $C_1 \cap C_3$  are obviously contradictory.

It is in fact the statistic for  $C_1 \cap C_3$  that causes most of the difficulty in Kyburg's example; as the process by which we arrive at it seems *prima facie* inconsistent, it can hardly be claimed that the difficulty sketched here constantly confronts us. Perhaps one could concoct a story as to how

much a statistic arose; more plausibly, one could change slightly the statistics in the example to avoid outright contradiction. In either event we will require further demonstrations that such examples are very common - a demonstration that transition to narrower reference classes constantly leads from knowledge of frequencies within narrow bounds around common points to knowledge of frequencies within wide bounds around more disparate points. Of course such cases can always be imagined and do, presumably, occur from time to time; however Kyburg's claim that such cases constantly arise now begins to look rather implausible.

A final point should be made about Kyburg's example; let us assume that we have statistics which indicate the proportion of  $G$  among the three equally narrow classes  $C_1 \cap C_2$ ,  $C_1 \cap C_3$ , and  $C_2 \cap C_3$  to be in the intervals  $(.70, .85)$ ,  $(0.75, 0.95)$ , and  $(0.75, 0.85)$  respectively. Here we may simply conclude that the probability of an  $n$  which belongs to  $C_1 \cap C_2 \cap C_3$  being  $G$  is in the interval  $(0.70, 0.95)$ . The basis for this conclusion is not that we expect the proportion of  $G$  in  $C_1 \cap C_2 \cap C_3$  to be in this interval (although such an expectation is not unreasonable) - it is that the three narrowest classes to which  $n$  belongs and for which adequate statistics are available all have frequencies within this rather broad interval. In Kyburg's example we are meant to conclude that an  $n$  which belongs to  $C_1 \cap C_3$  but not  $C_2$  has a probability in the interval  $(0.75, 0.95)$  of being  $G$ . If such a vague estimate of probability is acceptable, why should we not be satisfied in other cases with the only slightly more vague estimate of a

probability in the interval (0.70, 0.95)? For many practical purposes such a vague estimate of probability is all that we require and such an estimate can be thought of as a numerical equivalent of everyday statements in which we hold events to be quite probable, but not certain.

To treat such examples in this way is only to extend slightly a point made already. At the start of the discussion of (d), I noted that if we had numerous 'equally narrow' reference classes with a common limit, we could assign that limit as the unique probability for the individual event in question. Now in reality we must extrapolate to limits from observations in finite sequences and the resulting estimates will always be of limits within a certain interval, a certain specified degree of accuracy. For a set of equally narrow reference classes to have a common limit is then, in fact, for their limits to be expected to lie within a common interval. The unique probability assigned to an individual event on the basis of a 'common limit' will then only be a probability within a certain degree of accuracy - the degree of accuracy determined by the bounds of the common interval. It is only a small extension of this procedure to hold that events which belong to equally narrow reference classes whose limits are estimated to lie within different intervals are to be assigned unique probabilities within a degree of accuracy that includes all possible limits of the different classes. Such a proposal will only be of use if the different intervals within which the limits are expected are not very disparate but, I should think, this will usually be the case when there is the kind of statistical conflict Kyburg envisions. Thus in allowing ascriptions of



unique probabilities within a degree of accuracy we substantially increase the chances of assigning unique probabilities within such degrees of accuracy.

As we are more interested in analyzing the philosophical problems surrounding the principle of choosing the narrowest available reference class than in constructing an elaborate methodology, these somewhat sketchy remarks will suffice for our discussion of (d), for all we required was an indication of how we might maximize our chances of assigning unique probabilities. The above proposal to assign unique probabilities within a degree of accuracy coupled with a restriction to reference classes based on lawlike predicates should suffice for this purpose.

Turning to (c) above, it must be admitted that the argument in Chapter III relied heavily on the idea of a set of classes for which we can ascertain that individual experiments fulfill the defining characteristic prior to knowing the result in question, even if we have not done so. Explication of clause (2) in this respect is not an idle exercise in rigour, it is the first step in elucidating a concept of availability which is central to a wide variety of theories of probability. The problem involved in explicating clause (2) can be dealt with more easily if we distinguish two methods of forming reference classes: those which rely on mathematical functions on the numbering and order of success of the elements of a sequence and those which rely on empirical predicates. Mathematical functions only allow us to form reference classes narrower than one already given, e.g. the reference class consisting of every second element in the sequence of tosses of a coin. Empirical

predicates can determine reference classes directly, e.g. the class formed by the predicate 'x smokes 60 cigarettes a day', as well as classes narrower than an original one, for example, the sub-class of 40 year old men who smoke 60 cigarettes a day. Now, of course, it is clear that the mathematical function which selects every second element of a sequence yields a reference class acceptable or 'available' in the sense of clause (2); similarly it is clear that our empirical predicate based on cigarette consumption yields a reference class available in the sense of (2). But a solution to our problem (c) requires a general and exact characterization of each means of determining acceptable reference classes.

How should we set about giving such a characterization? An indication of one plausible approach can be found near the end of the previous Chapter where we saw that Salmon, building on suggestions of Reichenbach, proposed to reformulate the principle of choosing the narrowest available reference class in terms of choosing the broadest homogeneous reference class. Homogeneous reference classes were then defined in a manner similar to random classes. Although, as we saw, Salmon missed much of the significance of the relationship between randomness and probabilities for the single case, his suggested reformulation captured something of this relationship. Now that we are explicitly concerned with reformulating exactly the principle of choosing the narrowest available reference class, we might do well to follow up Salmon's proposal and employ the concept of randomness directly in our reformulation. The primary advantage in doing so is that we can thereby

bring latter day refinements of the concept of randomness to our aid in explicating clause (2). Salmon himself apparently overlooked this possibility - or was not interested in it - for the concept of randomness he briefly refers to in his reformulation is Von Mises' original concept, which, as we will soon see, is not very satisfactory for explication (2).

It is natural to begin an explication of (2) by delineating rigorously the set of mathematical functions which form reference classes acceptable under it. As a first, and very obvious step, we might identify this set of functions with those that select an infinity of elements  $x_i$  from an initial reference class  $x_1 \dots x_i \dots$ , where the rule of selection uses only the number  $i$  and the results  $x_1 \dots x_{i-1}$ . Mathematically the idea of a rule of selection based on the index number  $i$  and the results  $x_1 \dots x_{i-1}$  is usually expressed by a function on an infinite sequence of integers which encodes for each  $x_i$  the index number  $i$  and the results  $x_1 \dots x_{i-1}$ ; the values to this function determine which elements  $x_i$  of the original sequence are to be selected. Thus, the first step in explicating clause (2) consists simply in stipulating that the set of acceptable mathematical functions is the totality of all such selection functions on the sequence encoding the index number  $i$  and the results  $x_1 \dots x_{i-1}$ .

Unfortunately this proposal does not constitute any sort of advance for, contrary to the assumption up to now, the idea stemming from Von Mises of all selection functions on the sequence encoding the index number  $i$  and the results  $x_1 \dots x_{i-1}$  is known to be unsatisfactory for the

purpose intended.<sup>6</sup> So, as mentioned above, Von Mises original definition of randomness is of little use to us here.

Happily considerable research has been directed to finding a criterion similar in intent to Von Mises original one but more rigorous in detail; The most satisfactory result in this area is Church's well-known proposal in terms of effectively calculable functions on the sequence encoding the index number  $i$  and the results  $x_1 \dots x_{i-1}$ . This proposal can be adopted in a perfectly straightforward way to our present concern: the set of functions which form reference classes acceptable under clause (2) is stipulated to be the set of effectively calculable functions on the encoding sequence. More explicitly, a narrower reference class  $A_1$  derived from an original reference class  $A$  by a mathematical function  $\phi$  on an encoding sequence is 'available' for determining the probability of individual experiments  $x_n$  of  $A$  if and only if  $x_n \in A_1$  and  $\phi$  is effectively calculable.

There is a strong and, I think, interesting philosophical rationale for adopting such an explication of (2). The rationale here is suggested by Church's own justification for his proposal. Just before giving the mathematical formulation of his proposal Church stresses that gambling systems - the exclusion of which constitute randomness - should be defined in such a way that we have an effective procedure to determine what elements we are to gamble on.

To a player who would beat the wheel at roulette a system is unusable which corresponds to a mathematical function known to exist but not given by explicit definition; and even the explicit definition is of no use unless it provides a means of calculating the particular values of the function.



Church achieves the desired result by defining gambling systems in terms of general recursive, or effectively calculable, functions on an encoding sequence. The crucial point is that each effectively calculable function will not only determine an infinite sub-sequence  $A_1$  from an initial sequence  $A$ , but will also constitute a procedure by which in a finite time we can decide of any particular  $x_n$  of  $A$  whether it belongs to  $A_1$ . The use of a sequence encoding the index numbers  $i$  and the results  $x_1 \dots x_{i-1}$  insures that the procedure by which we decide this for a given  $x_n$  relies only on index number  $n$  and the results up to  $n-1$ .

It should thus be clear that by stipulating<sup>that</sup> the reference classes to which  $x_n$  belongs that are available under (2) are those formed by effectively calculable functions on an encoding sequence, we insure that there always exists a procedure for ascertaining (deciding) that  $x_n$  belongs to an available reference class. The procedure is the calculation of the values of our recursive function up to a point where we find whether or not  $x_n$  belongs to the class in question; moreover since the arguments to the recursive function are numbers from a sequence encoding only the index number  $i$  and the results  $x_1 \dots x_{i-1}$ , our procedure will never rely on the result  $x_n$  in question. This answers exactly to our needs in explicating (2) for we wish to explain the sense in which, without recourse to the result of  $x_n$ , we can ascertain that  $x_n$  belongs to a given reference class, even if we do not actually do so. On the proposal we are now considering the sense in which we can ascertain that  $x_n$  belongs to the reference class is that there exists a procedure which would yield this

conclusion if it were carried out. Here, as in other mathematical contexts, we equate the existence of a procedure that would yield a particular conclusion with the existence of a recursive or effectively calculable function that leads to that conclusion.

Thus the research done on randomness, culminating in Church's proposal, provides for mathematical functions the general and exact characterization required for a solution to (c). Moreover the argument so far suggests a solution to (c) for empirical predicates as well. The crucial point derived from Church's definition of randomness is that we can give a perfectly clear sense to the troublesome concept of 'being able to ascertain p' in terms of the existence of a procedure which, if carried out, would tell us that p. This suggests that we may delineate the empirical predicates that are acceptable under (2) by designating a certain finite set of predicates which correspond to extant scientific procedures, such as tests for time, spatial co-ordinates, temperature, cigarette consumption etc. As with recursive functions, the intuitive rationale for holding reference classes determined by these procedures to be available for measuring the probability of individual experiments is that we can always perform the relevant procedure and ascertain that the experiment belongs to the reference class, even if in reality we do not.

Of course not all predicates based on existing scientific procedures will figure in our finite list - we must confine ourselves to those procedures which do not involve knowledge of the result in question. To say this is not to define the intended class of predicates - a definition couched in such terms would be circular for we are attempting to give sense to the concept

of those predicates which can be known to apply to experiments prior to knowing the outcome of the experiment. Sense is fixed for this concept by explicitly defining a finite set of predicates as its extension and the above remark on avoiding use of the result in question is only intended to remind us of the basic motivation behind this definition.

The actual procedures designated as determining acceptable, or available, predicates will, of course, depend on the kind of phenomena we are interested in; moreover it will be researchers familiar with the field of study in question, rather than philosophers, who will be the most suited for specifying just what procedures exist for determining probability values. (The term 'field of study' like that of 'scientific procedure' may be understood loosely here - if we are concerned with horse racing, those familiar with the field of study will include informed punters, jockeys, trainers etc. The 'scientific' procedures they might specify would typically include those sufficient to determine a horse's blood line, general health, performance in training etc.)

In principle it should be possible to give an exhaustive specification of all the extant procedures appropriate for every field of study - the number of scientific experiments already carried out is obviously finite and this alone insures that the number of procedures used to determine the parameters in these experiments is finite. However, it would obviously be a burdensome task to carry out a specification of the appropriate procedures extant at this time and so we must confine ourselves simply to indicating that

such a specification would constitute for empirical predicates an explication of clause (2) analogous to that given above for mathematical functions in terms of effective calculability. In both cases the explication of the possibility of ascertainment referred to in (2) is given in terms of extant procedures that would so ascertain. For the case of mathematics the set of extant procedures will be the set of recursive functions. For the empirical case the set of extant procedures can only be given by (finite) enumeration.

It is worthwhile to note that the class of empirical predicates corresponding to scientific procedures extant at a particular time is similar to, but narrower than, the class of 'lawlike' predicates discussed earlier. The best known characterization of 'lawlike' predicates is Goodman's own in terms of entrenchment; roughly speaking 'lawlike' predicates are those that have become 'entrenched' by repeated use in the past. Obviously predicates corresponding to extant scientific procedures will be entrenched in Goodman's sense. Our class of predicates will be narrower than that of 'lawlike' predicates in virtue of excluding those which involve use of the result of the experiment in question.

This proposal to explicate (2), which I think suffices to resolve the difficulty expressed in (c), can be extended to provide part of an explication of the important concept of available evidence, and it is now convenient to examine some of the details of this extension. The similarity between the requirement of total evidence and our principle of choosing the narrowest available reference class has already been noted and is, in fact, rather obvious:



in using the frequency expected in a reference class  $A$  to measure the probability of an individual event  $x_n \in B$ , we are, of course, relying on the evidence that  $x_n$  belongs to  $A$ . Progressively narrower reference classes  $A_1, A_2$ , etc. to which  $x_n$  belongs obviously correspond to more and more detailed specifications of the conditions under which the experiment  $x_n$  is held. In using the narrowest reference class available to us, we in effect employ the most detailed (i.e. total) evidence available to us.

Furthermore it seems natural to say, on analogy to clause (2) for reference classes, that evidence available to us for an experiment is that which we can ascertain prior to knowing the outcome of an experiment. In light of the above considerations we can now give an exact characterization of what evidence might be so ascertained. Evidence which can be so ascertained is evidence which would be obtained - prior to knowing the result - if some extant procedure were carried out. The concept of extant procedure is then defined as above, that is, in terms of recursive functions on an encoding sequence when we are concerned with evidence about the number and position in order of success of an event in a sequence of events and in terms of a finite list of procedures for the case of 'empirical' predicates.

Such an explication of available evidence and available reference class leads to a highly objective criterion of availability - what is available to us is fixed quite objectively by the current state of science. Accordingly if we make our assignments of probability relative to such strongly defined available evidence (or reference class) we arrive at a conception of probability that will,

to some degree, satisfy those seeking an 'objective' account of probability.

Mellor<sup>7</sup>, for example, argues that frequency and logical relation theorists are not able to give a satisfactory account of 'plainly non-relational objective probability statements':

Their attempts founder on the problem of setting a non-arbitrary limit to the amount of evidence, or closeness of specification of the reference class, to be invoked in making the probability assignment.

The proposal considered here obviously does provide such a non-arbitrary limit and indeed carries out Mellor's subsequent suggestion:

The problem might be overcome, for example by an adequate explication of available evidence.

Later on I shall elaborate on the sense in which such an explication of available evidence leads to an objective conception of probability. More important for our present interest is that in giving an explication of available evidence (or reference class) which is sufficiently strong to capture some of the objective features of the concept of probability, we are led to a criterion demonstrably stronger than that found in actual practice. One instance of the kind of problem involved here has already been cited - the diseased horse of Chapter III. The general problem has been formulated under (b). In terms of the present discussion, the difficulty now is that available evidence is explicated in such a strong way that a person who fails to carry out any extant scientific procedure which yields relevant information at the appropriate time will fail to make use of all the available evidence. Similarly, of course, he will have failed to use the narrowest available reference class and, as our

example showed, this places unnaturally high demands on even the prudent researcher who wishes to use the narrowest available reference class.

Moreover, as we will see later, our explication of clause (2) for mathematical functions leads to analogous difficulties; for the moment, however, the problems arising from the explication for empirical predicates are sufficient for our attention.

We can easily multiply examples similar to the original one in which it seemed unsatisfactory to say information concerning a rare and undetected disease was available to the interested and informed punter. One drawn from Reichenbach highlights well the seriousness of the problems at issue here. Reichenbach remarks of a game of roulette that with

...precise observation of the initial velocity of the spinning ball, it should be possible to foretell with any degree of exactness where it will come to rest. It is only lack of technical ability that prevents us from equalling Laplace's superman. Once the velocity of the spinning ball has died down noticeably, such a computation comes into the scope of technical possibilities. This is well known to owners of gambling places, who stop such attempts by the croupier's call, "Rien ne va plus".<sup>8</sup>

The question for us is: what are we to take as the total evidence available just prior to the croupier's call? If this evidence - in relation to which we are to make our judgment of probability - includes highly precise specifications of the ball's velocity then, contrary to our ordinary assumption, the end result is not a matter of chance. If we accept Reichenbach's hypothesis that such precise observations are beyond our technical resources, we can

deny that this 'evidence' is available to us, for, by hypothesis, there exists no scientific procedure that would yield the information. But clearly now (and presumably also at the time Reichenbach wrote) numerous scientific procedures do exist which could determine with sufficient accuracy the velocity of the spinning ball in time to predict the result in question. Such procedures may only be able to be carried out with elaborate and bulky equipment but nonetheless they do exist.

Must we then conclude that evidence of the ball's initial velocity is available to us and, accordingly, that roulette is not truly a game of chance? From this conclusion it would be a very small step to conclude that a wide variety of events are actually determinate, but that, failing to consider all the available evidence, we mistakenly take them to be probabilistic. Such a conclusion - which Blackburn claims is inevitable on an objective conception of probability<sup>9</sup> - often seems lurking dangerously round the corner for various theories of probability and is one which should be avoided, I think most will agree, if at all possible.

It is perhaps all too obvious that we are on the verge of such a conclusion only because of the strength of our explication of available evidence or reference class. And of course it was the strength of our original vague formulation of clause (2) that first prompted our objection (b). It thus is apparent that we should give an explication of clause (2) weaker than the one already considered. To see the general form such an explication should take, we need only recall that our precept of using the total available evidence,



like that of using the narrowest available reference class, is only intended as a precept to guide our actions in the face of uncertainty. But as Ayer points out:

A methodological principle, considered simply as a guide to action, may be expected to be vague: so let us say that the principle is simply that we should always try to maximise the relevant evidence. In actual practice, this could be taken too far. There are times when it is rational to act on evidence which one knows to be incomplete: a general who refused to launch an attack until he had ascertained the position of every enemy soldier would not be very successful. There may be moral as well as practical limitations. If I am trying to pick a winner on a race-course, the principle of maximising evidence may be held to fall short of requiring me to attempt to bribe the stable boys.<sup>10</sup>

Besides the explicit point he makes, Ayer's remarks here illustrate well the variety of circumstances in which we seek to maximize evidence, and for us this is crucial. Every action in the face of uncertainty involves a decision made on imperfect knowledge, that is, on knowledge which is not sufficient to predict an outcome with certainty. Naturally in all such circumstances we desire to maximize our knowledge, that is, maximize the relevant evidence known to us. However, the circumstances in which we are required to make decisions with imperfect knowledge are many and varied. On the one hand we have occasions (rather like those facing the general) in which life and limb depend on the outcome of a trial for which complete knowledge is difficult or impossible. At the other extreme we find occasions in which few practical consequences follow and for which we impose artificial constraints on

our knowledge for the pleasure of testing our capacity for decision making under imperfect knowledge. Card games for small stakes with imperfect disclosure are clear examples of this. In between these extremes lie a whole host of circumstances in which imperfect knowledge figures in yet different ways - decisions on stock speculation are obviously of sufficient importance to require serious efforts at attaining the maximum possible information but, sensibly enough, many governments enforce special restrictions preventing the use of 'inside' information. In roulette the practical consequences may be little or great and the restrictions on information would be imposed in the end by casino owners.

With such a myriad of circumstances in which we either need or desire to act with less than perfect knowledge, it would be unreasonable to expect the unambiguous guide for action; in other words, any precept governing our actions in the face of imperfect knowledge will have to be sufficiently pliable to allow diverse application to very different circumstances. Since the precept to choose the total available evidence (or narrowest available reference class) is our guide to action in the face of imperfect knowledge, we must expect in actual practice a multiplicity, or family, of related criteria for applying these concepts, rather than a single clear criterion. The cause of our present difficulties now appears to be that the criterion of available evidence, or reference class, given by our first explication of clause (2) is appropriate for certain situations, but wholly inappropriate for others.

Our game of roulette provides a particularly clear illustration of this - if the total available evidence is defined to include any evidence that could be obtained at the right time by an extant scientific procedure, the ball's eventual resting place will be wholly determinate on the total available evidence. But of course it is rather implausible to think casinos would permit the fairly elaborate testing required to measure the crucial factor of the ball's initial velocity. And so, in a very real sense, such evidence is not available to actual gamblers. Indeed, as Reichenbach indicates, the croupier's call 'Rien na va plus' insures that the evidence available to gamblers in wagering is even less than can be gathered by the naked eye prior to the outcome of the wager.

Such restrictions on the evidence 'available' to us in various situations may, for convenience, all be termed 'practical'. But because of the complexity of our practice, these practical restrictions can not be codified in a way that yields one clear criterion of what is available evidence, or what is an available reference class. If this view is correct, we have it on Wittgenstein's authority that any exhaustive codification of what in varying circumstances we ordinarily regard as available evidence (or reference class) would be exceedingly complex. Accordingly I will only try to indicate briefly the broad outlines of our ordinary concept of available evidence (similar remarks, of course, will apply to the concept of available reference class).

At one end of a continuum of criteria for what constitutes available evidence, there is the criterion we have in mind when we judge what evidence is available to a scientist engaged in research proper. Unlike the

card player who cheats if he tries to obtain all possible knowledge, a scientist engaged in research is expected to take all possible steps to maximize his knowledge. Accordingly, our formulation of a criterion for the evidence ordinarily considered 'available' to a scientist should be as strong as possible. The initial explication of available evidence for the case of empirical predicates, in terms of information resulting from the implementation of extant scientific procedures suffices here and, indeed, was proposed with this purpose in mind. A scientist who wishes to make probability judgments on the basis of all the evidence available to him therefore must ensure that he has not overlooked any relevant evidence ascertainable by an extant procedure of his science, for such evidence is 'available' to him. This seems a wholly satisfactory conclusion, indeed the only possible one. In our example of a horse with a rare but undetected disease, it seemed inappropriate to say that evidence of this disease was available to the average punter, as he could not be expected to obtain it. But a scientist attempting to confirm a physiological theory of equine motor activity would be expected to test for and uncover this disease if, for example, he were trying to explain why the horse failed to fulfill a prediction of his theory. Conversely, when attempting to predict the horse's motor activity on the basis of his theory, he would be expected to carry out this and any other relevant test and so it is plausible to characterize such information as available to him.

It was remarked earlier that adopting an explication of available evidence along these lines led to an objective conception of probability. The reason for this should now be apparent: objective judgments are those which



conform to the standards set by the current state of science and so probability judgments based on all the evidence which can be obtained by extant scientific procedures will be, in an important sense, objective. In light of this, I will, in what follows, speak of the objective criterion, or objective sense, of 'available evidence'; by 'available evidence' in this objective sense I will mean that evidence which can be known by implementing any relevant extant scientific procedure.

At the furthest remove from such an objective sense, or criterion, of available evidence, some have defined the evidence available to a person as that of which he is consciously aware. Although this criterion is so weak as to collapse entirely the distinction between available evidence and known evidence, it is worth some attention: If we understand available evidence in this weak sense the requirement of total evidence reduces to the very weak, but not implausible, requirement that probability judgments made by a person must be relative to the total evidence known to him at that time he makes the judgment. Carnap is the most eminent writer who understands available evidence in this weak sense:

Suppose that inductive logic supplies a simple result of the form ' $c(h, e) = r$ ', where  $h$  and  $e$  are two given sentences and  $r$  is a real number. How is the result to be applied to a given knowledge situation? The question is answered by the following rule, which is not a rule of inductive logic but of the methodology of induction:

1. If  $e$  expresses the total knowledge of  $X$  at the time  $t$ , that is to say, his total knowledge of the results of his observations, then  $X$  is justified at this time to believe  $h$  to degree  $r$  and hence to bet on  $h$  with a betting quotient not higher than  $r$ .

One of the decisive points in this rule is the fact that it lays down the following stipulation:

2. Requirement of total evidence: in the application of inductive logic to a given knowledge situation, the total evidence available must be taken as a basis for determining the degree of confirmation...<sup>11</sup> [My underlining]

It may well be that there are circumstances in which such a weak requirement is the most appropriate one for judgments of probability. But of course this weak form of the requirement of total evidence can not do in general: as Ayer points out<sup>12</sup>, with such a requirement we need never look for new evidence in making probability judgments.

It is also worthwhile to note that acceptance of Carnap's weak requirement of total evidence leads to a form of subjectivism in the application of his theory to obtain the unique probability values we so often believe in and use as the basis for action. Of course, as Carnap repeatedly stressed his theory of probability is objective in the sense that the degree of confirmation given to a hypothesis  $h$  by evidence  $e$  is independent of any particular person's judgments.<sup>13</sup> However, if the requirement of total evidence imposed is the weak one suggested by Carnap, the application of his theory of confirmation to a given situation to determine 'unique' results will actually give results that vary subjectively from person to person, depending on what evidence that person knows. The degree of probability<sub>1</sub> determined by  $X$  in a given situation will be the (objective) degree of confirmation  $c$  given to the hypothesis  $h$  by the total evidence  $e$  known to  $X$  at that time. Thus the value of  $c$  for

h decided on for the purposes of action depends on the subjective matter of how much evidence X actually knows and will be different for people who know different amounts of evidence.

Whether or not this is a serious flaw in Carnap's methodology need not detain us in this chapter; his remarks here are primarily of interest to us in setting a lower bound for the range of criteria we employ in determining what evidence is available to us. The upper bound has already been fixed in terms of that which can be obtained by implementing extant scientific procedures - available evidence in the so-called objective sense of that word. When, as with Carnap, the evidence available to a person is understood as that which happens to be known to him, I will say that 'available evidence' is being understood in a weak subjective sense. I myself am in some doubt if there are any circumstances in which we actually understand 'available evidence' in such a weak subjective sense, but, obviously enough, Carnap believes there are and I, for the moment, am content to take him at his word.

In between such extreme circumstances in which available evidence would be understood in, respectively, the objective and subjective sense, there is a middle ground comprising a large variety of circumstances in which the evidence ordinarily deemed available to us is limited by what I have called 'practical restrictions'. Obviously in each of these circumstances the evidence deemed available will be some proper sub-set of that which could be obtained by the totality of extant scientific procedures, e.g. the evidence 'available' in a casino is a proper sub-set of that available in a laboratory;

similarly if the evidence which we regard as available in these circumstances coincides with that known by a particular person, this will only be due to his diligence in acquiring evidence, for the evidence we regard as available to him will not, as a matter of definition, be just that known to him.

This latter consideration suggests one feature common to many circumstances in which probability judgments are made relative to a body of evidence deemed the total available. Even when the criterion we have in mind for 'available evidence' is weaker than the objective one appropriate to the practice of science proper, it is primarily an inter-subjective criterion. By this I mean that evidence deemed available to one person in a given situation will ordinarily be thought available to all in a similar situation. Honest gambling houses make provision to ensure no privileged information is accessible to a minority of players and, as we saw, stock markets often prohibit participation of those with 'inside' information. Even card games in which privileged access to his own hand is given to each player are only a trivial exception to this - all information other than that of his own hand is available to each player on an equal basis.

Perhaps there are other features common to the evidence deemed available for those circumstances which fall between our two extremes. This writer can not readily think of any and, as we are dealing with extremely practical matters, the task of elucidating such common features (if they exist) would be difficult. I will therefore adopt this somewhat rough - and not necessarily exhaustive - distinction between objective, subjective and inter-



subjective criteria of available evidence and conclude this discussion of (b) with a few general remarks about these different kinds of criteria. It should be obvious that these different criteria for available evidence also constitute different criteria for reference classes deemed available in the sense intended by clause (2). Indeed the entire discussion above could have been (and in places was) expressed in terms of what reference classes are plausibly thought of as available to us for determining probability values. Conversely, although (b) initially was expressed as an objection to construing a wide range of classes as available for determining probabilities, it is obviously equivalent to an objection to construing the parallel range of evidence available. The reply to (b) - in either form - which has emerged above is that for a wide range of phenomena no one set of scientific procedures can be delineated to provide a plausible explication of what is meant by available evidence or available reference class. Evidence 'available' to a scientist in his laboratory is not the same as that 'available' to a gambler playing roulette in a casino. This point, to some degree implicit in Ayer's remarks on the vagueness of our methodological principle, was expressed above by saying that the concept of available evidence (and *pari passu* available reference class) had a multiplicity, or family, of criteria of application in our actual practice.

Accepting such a viewpoint allows us to give a perspicuous<sup>u</sup> solution to a serious problem encountered, and briefly analyzed, above. Adopting the strongest criterion of available evidence, the objective, we saw how ordinary games of 'chance', such as roulette, could easily appear deterministic - if the available evidence means that which a scientist using

all extant procedures could obtain, the outcome in roulette is determinate on the total available evidence well before any wagers need be made. A similar point could be made about many other apparently probabilistic matters and we can thus easily be led to the unhappy conclusion that for many cases ordinarily thought matters of chance the 'actual probability is something near 1 or 0, but that normally we do not know which', as Blackburn puts it<sup>14</sup>. This is of course a perennial philosophical difficulty - Hacking, for example, discussed it nearly a decade before Blackburn.<sup>15</sup>

However, in the example of roulette the reasoning to this conclusion was seen to involve the application of a criterion of available evidence, the objective, to situations in which this was demonstrably inappropriate. Again invoking the name of Wittgenstein we can now generally diagnose this philosophical difficulty as arising from a disregard of the multiplicity of criteria found in our complex practice. Cases for which the objective criterion of available evidence is appropriate - the practice of science proper - are fastened on to and tacitly treated as paradigmatic. Application of this criterion to very different cases - ones in which our knowledge is deliberately limited for the sake of sport - leads us into the quandry: surely roulette is a game of chance, yet surely it is deterministic. The way out of the quandry is to see it as arising from treating as paradigmatic a criterion for available evidence which in fact is only appropriate to a limited range of cases.

Our rough classification here of objective, subjective and inter-subjective criteria of available evidence only pertains to evidence based on what I have called empirical predicates. This classification arose from the realization that the explication of (2) which sufficed to resolve (c) for the empirical case would have to be refined further in light of (b) - in particular the finite list of extant scientific procedures which first fixed the sense of (2) constituted an objective criterion for available evidence that codified our practice in scientific research proper, but was too strong for other cases. Sub-sets of this finite set would represent the evidence ordinarily deemed available in these other cases because, to put it roughly, in different contexts different kinds of evidence is thought available for determining probabilities. But the question remains whether there is one set of mathematical functions which can plausibly be thought to determine available evidence or reference classes in every context. The issues involved here are most easily expressed in terms of the criteria employed in holding a reference class, rather than evidence, available and so I will return to explicit discussion of the reference classes which are to be held available in the sense of clause (2).

Initially, following Church, it was proposed that the set  $R$  of recursive functions on an encoding sequence fixed the set of functions acceptable under (2). Although this was a vigorous - and fruitful - suggestion, it now must be admitted that it was at best an idealized, and therefore unrealistic, proposal. In essence our problem (b) was that in many contexts there were a number of extant scientific procedures that we could not expect even the prudent non scientist to carry out; accordingly, we could not reasonably

hold every reference class based on such a procedure to be available to us.

But if we can not regard every reference class formed by one of a finite number of relevant extant scientific procedures to be available to us in every context, we can hardly regard every reference class formed by a denumerable infinity of recursive functions to be available to us in every context.

The difficulties involved here have been covered up so far by our assumption (1') that any infinite sequence we are presented with is one for which we know all the limits in its sub-sequences. Although this was a convenient (and common) idealization, the reality is that we can only observe frequencies in finite sequences and their sub-sequences. All 'knowledge' of limits must be gained by estimation and extrapolation from observation of such finite frequencies.

It is to this reality of limits estimated on the basis of frequencies observed in finite sequences that any methodological principle for determining unique probabilities must apply and, unfortunately, to this reality the proposed explication of (2) for mathematical functions is patently inappropriate. To see this suppose we observe the frequency  $p$  for an attribute  $B$  in a finite sequence  $a$ ; on the basis of this observation we expect  $p$  as the limit of  $B$  if the sequence  $a$  were extended indefinitely to form an infinite sequence  $A$ . (Here and throughout this discussion I assume for simplicity that we use the so-called straight rule of induction for estimating limits on the basis of finite observations). The next experiment that occurs in the indefinite extension of  $a$  is, say,  $x_n$  and we desire a unique probability value for the experiment  $x_n$



turning out to have the property B. Before we can use the expected limit  $p$  in A as this value, we must determine that  $x_n$  does not belong to a narrower available reference class with a different limit. As we are now concerned with the explication of (2) for mathematical functions, we may disregard the possibility that A can be sub-divided by use of empirical predicates. (Similarly we may, for simplicity, ignore the possibility discussed under (d) - that there might exist other equally narrower classes to which  $x_n$  belongs.) Thus before we can regard the indefinite extension A of a as the narrowest available reference class for determining the probability of  $x_n$  we need - for our present purposes - only ascertain that A can not be sub-divided by mathematical rule into some infinite sequence  $A_1$  to which  $x_n$  belongs and for which we expect a limit  $p_1 \neq p$ . Since - and this is the important point - we only have the finite sequence a from which to extrapolate to limits, the limit  $p_1$  in  $A_1$  will be estimated as equal to the frequency found in the finite sub-sequence selected from a by the same mathematical rule which selects  $A_1$  from A. I will call this finite sequence  $a_1$ .

Thus, for example, if we have tossed a coin 1000 times (the finite sequence a) with roughly 500 heads, a unique probability of  $\frac{1}{2}$  can be assigned the next toss  $x_n$  by the principle of choosing the narrowest available reference class, if we believe (1) that indefinitely repeated tosses of this coin under the same conditions (the infinite sequence A) would have a limit (p) of  $\frac{1}{2}$  for heads and (2) that there is no infinite sub-sequence among these repeated tosses A which would have a different limit for heads and to which  $x_n$  also belongs. If, to take the classic case, every odd numbered toss in the first 1000

(the sub-sequence  $a_1$ ) turned out to be heads, we could expect the limit ( $p_1$ ) of heads in the infinite sub-sequence of all odd tosses of this coin ( $A_1$ ) to be 1. In such a case we would have to assign the probability 1 to the 1001st toss ( $x_n$ ) by the principle of choosing the narrowest available reference class.

Now, to return to the general problem, of course we need not (and could not) determine that  $x_n$  belongs to no infinite subsequence  $A_1$  such that the expected limit  $p_1 \neq p$ ; we are only interested in sub-divisions of our indefinitely extended sequence  $A$  which are based on functions that fulfill some version of clause (2) e.g., the sequence of every odd toss in the above example. The problem is that the explication proposed earlier for (2) in terms of recursive functions  $R$  on an encoding sequence clearly will not now do - it is absurd to think of someone attempting to consider each recursive function  $R_1$  which selects a sub-sequence  $A_1$  of the indefinite extension  $A$  of a finite sequence  $a$  and then checking to see if the expected limit  $p_1$  based on the frequency in  $a_1$  equals the expected limit  $p$  in  $A$  based on the frequency in  $a$ . Among other problems, the cardinality of the set of such recursive functions is infinite and so, of course, we can not actually consider every recursive function.

The basic question to ask here is what in general would constitute a satisfactory delineation of the set of mathematical functions that form available reference classes, the set acceptable under (2). Abstracting from the point just made - in a somewhat unhappy psychological mode of expression at that - we may give a simple answer: for the principle of

choosing the narrowest available reference class to have genuine practical utility, we must be able to effectively (i.e. recursively) decide if the indefinite extension  $A$  of an observed finite sequence  $a$  will be the narrowest reference class available for a given experiment  $x_n$  or if, instead, there exists for  $x_n$  a mathematical function forming a sub-division  $A_1$  of  $A$  such that its expected limit is different from that expected in  $A$ . If such an  $A_1$  exists it will form a narrower available reference class than  $A$ .

Let me hasten to add here that the question for which we want an effective decision procedure is not whether, if extended indefinitely, our finite sequence  $a$  with observed frequency  $p$  will actually form an infinite sequence  $A$  with limit  $p$  which is the narrowest reference class available for its constituent experiments. As we can never conclusively verify that some sequence has a genuine limit - the results of subsequent experiments can always contradict the assumption of a given limit - we could not even have a decision procedure for the question of whether the indefinite extension of some finite sequence actually possesses a limit, much less whether it forms the narrowest available reference class. What we want is a decision procedure for the question of whether the infinite class  $A$  we envision at a time  $t$  as the indefinite extension of our finite sequence  $a$  can be regarded at  $t$  as the narrowest available reference class. If the finite sequence  $a$  is actually extended by carrying out more trials, all sorts of things can happen that prevent us from regarding it after  $t$  as the narrowest reference class available for any of its constituent experiments - a very obvious pattern might develop which

had not obtained before  $t$  etc. Obviously such eventualities can not be catered for at time  $t$ , so what we require is a procedure for telling us if the finite sequence observed at time  $t$  has already exhibited patterns which can be expected to continue if the sequence is later extended. If such a pattern has already occurred by  $t$ , as for example with a finite sequence of tosses of a coin alternating between heads and tails, we can not at  $t$  plausibly think that the indefinite extension of this sequence will be the narrowest reference class available for the individual tosses. Thus above, and in what follows, when I speak of a procedure for effectively deciding if the indefinite extension of a finite sequence is the narrowest reference class available, it must be understood that what I intend is a procedure which decides if a finite sequence can be regarded at the time of its observation as one which, if indefinitely extended, would form the narrowest available reference class.

In demanding a procedure which allows us to effectively decide if the extension of a given finite sequence is to be regarded as the narrowest reference class available, we are only giving a precise formulation to our need for a methodological principle for determining unique probabilities that we can be sure can be applied in practice. To be more exact, we are demanding a methodological principle that we can be sure is applicable if other things are equal, that is, if there are not numerous equally narrow reference classes with different limits or empirical procedures of sub-division to be considered. Since as indicated above, these considerations are irrelevant to our present concern, we may here ignore them and conveniently express the adequacy condition for the delineation of the mathematical functions acceptable



under (2) by saying that it should lead to an effective procedure for deciding if the indefinite extension of a given finite sequence is to be regarded as the narrowest available reference class.

Needless to say, if the set of mathematical functions which form reference classes available under (2) is delineated in a way that leads to an effective procedure for deciding this question, doubts of the kind expressed in (b) for our methodological principle are resolved for the case of mathematical functions. The researcher who wishes to know if the infinite extension  $A$  of a finite sequence  $a$  is the narrowest available reference class need only carry out a clearly defined procedure, for there will exist a determinate procedure which settles this question in a finite time. Conversely it seems clear that any specification of mathematical functions in explicating (2) that does not lead to such an effective procedure will be undesirable because there will be no assurance that we will be able to say if the indefinite extension of an arbitrarily given finite sequence constitutes the narrowest reference class available.

Unfortunately, as the reader may have already perceived, our initial explication of (2) in terms of recursive functions is effective in this sense, but for a most unwelcome reason. We have a finite sequence  $a$ , this we imagine extended indefinitely to form a reference class  $A$  which includes the next experiment  $x_n$ . Now whatever the frequency of  $B$  in  $a$  other than 0 or 1 (and hence whatever expected limit in  $A$  other than 0 or 1),

we can be sure that there exists some recursive function  $R_1$  which selects a sub-sequence  $A_1$  of  $A$  to which  $x_n$  belongs and for which all the experiments  $a_1$  selected by  $R_1$  from  $a$  have the property  $B$ . The expected limit of  $B$  in  $A_1$  would then be 1 and  $A_1$  then would be a narrower reference class than  $A$  which, moreover, is available to us for measuring the probability of  $x_n$  under the proposed explication of (2).

The problem raised in the above paragraph can be expressed perspicuously if we note, as the reader may have done already, that our use of Church's definition of randomness in explication (2) entails that the problem of deciding whether a given infinite sequence can be subdivided to form reference classes narrower than that sequence and yet available under (2), is equivalent to the question whether or not the sequence is random in Church's sense.

The problem cited in the last paragraph really stems from the well-known difficulty that the Church - Von Mises definition of randomness can not be extended to apply to finite sequences. If one were to define a finite sequence as random if and only if it contained no finite sub-sequence determined by a recursive function which had a different frequency of the attribute in question, one finds that the only finite random sequences are those with 0 or 1 as the frequency for the attribute. Any other finite sequence  $a$  will have numerous sub-sequences  $a_1$  picked out by some recursive function  $R_1$ , where all elements of  $a_1$  have the attribute in question. If we use Church's definition of randomness for our explication of (2), it also follows that no finite sequence

a with a frequency other than 1 or 0 can be the initial segment of an infinite class  $A$  which could be the narrowest reference class available for all its constituent experiments. Every such finite sequence  $a$  can always be sub-divided by a recursive function into a sub-sequence  $a_1$  with a frequency of 1. Since every sub-division carried out by a recursive function is supposed to yield an available reference class narrower than  $A$  itself, the infinite extension  $A_1$  of  $a_1$  will be an available reference class and one with an expected limit of 1. This was exactly the problem noted in the previous paragraph.

To avoid this difficulty we obviously require a definition of randomness applicable in an interesting way to finite sequences. Happily considerable research has been done on this problem, particularly by Kolmogorov and Martin-Lof<sup>16</sup>. In essence they propose to define a finite sequence as random if and only if the degree of complexity of the shortest programme required to compute the sequence is nearly as great as the complexity of the sequence itself. The complexity of a sequence and the programme used to compute it are defined relative to a particular language for automata and a particular calculating device, say a Turing machine. Roughly speaking, the complexity of a sequence and the programme used to compute it will correspond to their length; thus a random sequence is one that can only be produced by programmes nearly as long as itself. The basic idea here is that finite sequences which can be produced by programmes much simpler, i.e. shorter, than themselves are plausibly thought of as embodying

genuine regularities; those that can only be produced by programmes as 'complex' or long as themselves only involve artificial regularities that can be found for any finite sequence. The latter complex sequences can be regarded as random, while the former can not be.

The technical details of this definition are now widely accessible<sup>17</sup> and so need not detain us here. What is important for us is the fact that this proposal can be adopted to yield an explication of clause (2) that avoids the conclusion that only a finite sequence with frequency of 0 or 1 can be taken to be the initial segment of sequence that would form the narrowest available reference class. Again assume we observe a finite sequence  $a$  with a frequency  $p$  of  $B$ . As before to determine the probability of the next experiment  $x_n$  being  $B$  we imagine  $a$  extended to form the infinite reference class  $A$ ;  $p$  will then be the expected limit of relative frequency of  $B$  in  $A$ . Disregarding the possibility of sub-division by empirical predicates and other equally narrow reference classes, we may now say we are entitled to treat  $A$  as the narrowest reference class available for  $x_n$  if and only if  $a$  is random on the basis of the complexity definition given by Kolmogorov and Martin-Lof. If  $a$  is random in this sense the sub-divisions  $a_i$  of  $a$  that do exist with different observed frequencies will not be regarded as sequences which if extended indefinitely would yield new narrower available reference classes with a limit different from that in  $A$ . The intuitive rationale for this is that the sub-divisions  $a_i$  that have different frequencies than those found in a finite random sequence  $a$  will not arise from genuine regularities which



could be expected to continue if more trials were made, rather they will be the outcome of 'complex' and thus artificial selections on  $a$ .

The only remaining problem here is that already raised concerning the effectiveness of our concept of available reference class. As we saw, what we really required for the case of reference classes formed by mathematical rules of selection was an effective procedure for deciding whether the indefinite extension of a finite sequence  $a$  observed at time  $t$  can be regarded at  $t$  as the narrowest available reference class. As we have just now proposed to identify those finite sequences whose extension can be regarded as the narrowest available reference class with those that are random on the Kolmogorov - Martin-Lof definition, it follows directly that our new proposal will be effective in the desired sense if and only if the Kolmogorov - Martin-Lof definition of randomness is effective. Unfortunately, as is widely known, this definition is non-effective; that is, there does not exist a recursive procedure for determining whether a given finite sequence fulfills this definition of randomness.

Although in light of (b) this must count against the explication of (2) in terms of the Kolmogorov - Martin-Lof proposal, it puts our remaining problem into sharp focus. What is required for a thoroughly satisfactory explication of (2) is an effective definition of randomness applicable to finite sequences. If we say that the indefinite extension of a finite random sequence can be regarded as the narrowest reference class available for its constituent individual experiments, an effective definition of randomness allows us to decide

effectively if the indefinite extension of a particular finite sequence constitutes the narrowest reference class available for its constituent experiments. The task of finding such an effective definition of randomness applicable to finite sequences is a highly technical matter, regrettably beyond the scope of this thesis. It is, of course, widely appreciated that such a definition is desirable<sup>18</sup>, although (to this writer's knowledge) no such account has yet been given. We must therefore conclude our discussion of (b) with the observation that for the case of mathematical functions, we have yet to find a thoroughly satisfactory explication of (2).

We are left now with our problem (a) which, happily, can be dealt with more expeditiously than our previous problems (c) - (b). In Chapter III two formulations were given to our methodological principle, one in terms of clause (1), the other in terms of (1'). The discussion in Chapter III was, on the whole, carried out under the assumption expressed in (1') that we are given infinite sequences for which we know all the relevant limits. This assumption was appropriate for Chapter III as our main concern there was with Von Mises requirement of randomness and, of course, his definition of randomness was given on the assumption expressed in (1') that we did know all the relevant limits in infinite sequences. However, no matter what the merits of adopting the assumption (1') in discussing Von Mises' requirement of randomness, it is clear that any satisfactory formulation of the principle of choosing the narrowest available reference class must eventually be given in terms of clause (1) rather than (1').

Already several times in this chapter we have had to remind ourselves that in reality we are never presented with infinite sequences with known limits, but rather may only extrapolate to such limits on the basis of observation of finite sequences. The problem (a) that remains is to give a general account of the procedures by which we extrapolate from observed frequencies in finite sequences to expected limits in infinite classes.

Regrettably, but perhaps not surprisingly, no such account will be offered here. The precise specification of procedures by which such extrapolation may plausibly be carried out is a problem in statistics or, more accurately, the foundation of statistics. The question of whether any procedure of extrapolation from past observation into the future can ever be justified is, of course, just the problem of induction. Although there can be no clear dividing lines between the study of probability theory, induction and the foundations of statistics, I will here avoid as far as possible the knotty problems that have been so widely discussed in the latter two fields. To avoid lengthy discussion of these problems I will simply - or perhaps simplistically - accept as valid the so-called straight rule of induction, that is, the principle that we may expect the limit of frequency in the indefinite future extension of a sequence to match closely the frequency already observed in the finite initial segment.

Reichenbach, among probability theorists, is this rule's most ardent proponent, offering a lengthy and famous argument justifying its use<sup>19</sup>. Salmon offers a searching analysis of Reichenbach's argument

here and also indicates something of the relevant recent philosophical literature surrounding the straight rule of induction<sup>20</sup>.

Among statisticians, Fisher's influential principle of maximum likelihood can readily be shown to lead to the straight rule of induction when the principle is used for extrapolation to infinite sequences<sup>21</sup>. Hacking has offered a perspicuous analysis of a wide variety of statistical techniques in the hopes of showing the principle of maximum likelihood as fundamental for all statistical inference.<sup>22</sup> Carnap, on the other hand, has argued convincingly that the straight rule of induction - and thus Fisher's principle of maximum likelihood when applied to infinite sequences - is at best a convenient approximation for a continuum of more plausible methods.<sup>23</sup>

Since we are not here concerned with the foundations of statistics or inductive logic, we will take our cue from Carnap and resolve our last problem (a) by accepting the straight rule as the simplest and most convenient procedure of inductive extrapolation from finite data. In this chapter I already assumed the validity of this rule in the discussion of extrapolation from finite sequences and will, in what follows, again assume its validity wherever appropriate.

With this as a somewhat cursory consideration of our last problem (a), let me now summarize the findings of this chapter. We began by noting five prime facie difficulties for the methodological principle of choosing the narrowest available reference class as formulated in Chapter III.



Since the argument found there depended heavily on adopting this principle in one form or another, it was necessary to see if any of these difficulties were insurmountable. As the principle of choosing the narrowest available reference class has played an important role in many discussions of probability, the difficulties (a) - (e) were not merely problems for the argument of Chapter III but represented, in varying degrees, substantial issues of intrinsic interest. With regard to problem (e), we were able to reply to Ayer's objection on the basis of remarks made by Salmon and Reichenbach. In considering the problem raised in (d), we found that doubts over the efficacy of our principle in determining unique probabilities were less serious than some (e.g. Kyburg) have maintained. To deal properly with (c) we had to offer a rigorous explication of the concept of available reference class, which in turn led to an analogous explication of our concept of available evidence. Given the prominence attached to this concept in a wide variety of theories of probability, the findings here are, hopefully, of some general interest.

The explication initially offered of the concept of available evidence or reference class later had to be refined in light of (b). Indeed it turned out that for the case of empirical predicates, no one explicans would do for the explicandum figured prominently in a wide variety of different and complex practices. Inattention to the multiplicity of criteria found in the actual employment of the concept of available evidence or reference class was, moreover, diagnosed as the cause of one common and serious quandary over the concept of probability - the dilemma that for many matters ordinarily thought probabilistic, the probability of an event appeared to be nothing other

than our negligent ignorance in the face of its actually deterministic character.

The problem of delineating precisely - and yet usefully - the set of mathematical functions that form 'available' reference classes was given a partial solution. For the idealized case of infinite classes with known limits, we could adopt Church's definition of randomness and identify these functions with recursive functions. For the actual case of finite sequences, we adopted the Kolmogorov - Martin-Lof definition of randomness.

The only problem left outstanding in this discussion was that our proposed explication of (2) for mathematical functions left us without an effective procedure for determining when a particular finite sequence was the initial segment of a reference class that would be the narrowest available. This was seen to involve a technical mathematical problem and one that has already received considerable attention.

## Footnotes for Chapter IV

1. Ayer, A.J., op. cit. p. 52.
2. Gillies, D.A., An Objective Theory of Probability, London 1973, p. 156.
3. Salmon, W.C., op. cit. p. 96.
4. Hempel, C.G., Maximal Specificity and Lawlikeness in Probabilistic Explanation, Phil. Sci. vol. 35, 1968, pp. 116-33.
5. Kyburg, H., More on Maximal Specificity, Phil Sci. vol. 37, 1970, pp. 295-300.
6. Church, A., op. cit.,
7. Mellor, D., op. cit. p. 53.
8. Reichenbach, H., op. cit., p. 150.
9. Blackburn, S.W., Reason and Prediction, Cambridge 1973, pp. 106 ff.
10. Ayer, A.J., op. cit., pp. 56-57.
11. Carnap, R., op. cit., p. 211.
12. Ayer, A.J., op. cit., pp. 56-57.
13. Carnap, R., op. cit. Section 11.
14. Blackburn, S.W., op. cit., p. 106.
15. Hacking, I., Logic of Statistical Inference, Cambridge 1965, p. 24
16. Martin-Lof, P., The Definition of Random Sequences, Information and Control, vol. 9, 1966, pp. 602-19.
17. See Chaitin, G.J., Randomness and Mathematical Proof, Sci. Amer. vol. 232 no. 5 1975, pp. 47-52 for a good popular account of this definition, also Fine, T., Theories of Probability, New York 1974.
18. Fine, T., op. cit., p. 150

19. Reichenbach, H., op. cit., Section 87.
20. Salmon, W.C., op. cit., pp. 96-108.
21. Carnap, R., The Continuum of Inductive Methods, Chicago 1952, pp. 43-44.
22. Hacking, I., op. cit.
23. Carnap, R., CI, pp. 14-18.



## CHAPTER V

### Two Concepts of Probability

It has become the fashion in probability theory to differentiate a number of distinct concepts of probability. The most influential proponent of such a view is of course Carnap - his view that there are two concepts of probability is now a familiar part of probability theory. Recently Hacking, in his book on the emergence of probability, has accepted what I will call Carnap's thesis and, further, has argued that such a 'duality' has been implicit in usage of the term 'probability' ever since its 'emergence' around 1660. Moreover, he claims that the distinction between probability<sub>1</sub>, which he calls an 'epistemic' concept, and probability<sub>2</sub>, the 'aleatory' concept, was explicitly recognized by philosophers as early as 1785, in particular in the works of Condorcet<sup>1</sup>. Other writers go further than Hacking and Carnap and distinguish more than two concepts of probability. Mackie, who provides the most extended recent discussion of various concepts of probability, considers Carnap's thesis and concludes that there are 'at least five fairly distinguishable probability concepts with sub-divisions within several of them, as well as links between them'.<sup>2</sup>

The consequence of such a multiplication of concepts is, on occasion, somewhat unnerving. Hacking in his earlier book on the foundation of statistics remarks:

In the course of this essay we shall need to refer to practically everything which the word 'probability' has ever been used to refer to; to avoid confusion it is better to abandon the word altogether and use other expressions which have not yet been made so equivocal.<sup>3</sup>

Certainly probability theory has fallen into a sorry state if usage of its fundamental term has become so equivocal as to recommend its wholesale abandonment.

Such a terminological anomaly, moreover, points to a serious philosophical issue. If, using Occam's razor, we are to trim the growth of entities employed in philosophical analysis, an appeal to a multiplicity of concepts of probability can only be justified if unavoidable; thus we must in good conscience make every effort we can to reduce to a minimum the number of concepts used to account for our intuitions on probability.

That there may exist just one single concept of probability underlying an apparently diverse usage has already been suggested by Gillies<sup>4</sup>, Ayers<sup>5</sup> and Mellor<sup>6</sup>. Given the proliferation of theories of probability claiming to elucidate distinct concepts of probability it would be a very ambitious, perhaps even fool hardy, person who attempted to demonstrate in detail that all these concepts reduce to one. Accordingly I will here consider only a few prominent theories of probability to see to what extent, if any, we may locate a common concept amidst these apparently very different theories. In this chapter I will be concerned with the logical relation theories of Keynes, Jeffreys and Carnap. In particular I will try to show how, contrary

to Carnap's thesis, it is possible to identify one basic concept underlying the epistemic and statistical concepts of probability<sub>1</sub> and probability<sub>2</sub> originally distinguished by Carnap. The next chapters of the thesis will be devoted to a study of the classical theory of probability and its notorious companion, the principle of indifference.

Carnap's theory of probability<sub>1</sub> is of course the most sophisticated logical relation theory and so is the one with which I will be primarily concerned. On the face of it, Carnap's theory of probability<sub>1</sub> is, as he himself claimed, very different from any frequency theory of probability. Probability<sub>1</sub> is defined as a quantitative relation between propositions holding solely in terms of logical, or a priori, considerations; whereas, on any frequency theory of probability, probabilities will be assigned on the basis of factual considerations of long term frequencies. Secondly, probability<sub>1</sub> judgments are determined by measures of the 'range' of two sentences about our individual events, one of which constitutes the evidence for the other. Here the contrast with the standard frequency theory, which Carnap meant when he spoke of probability<sub>2</sub>, is even greater. First of all, even if the standard frequency is formulated as a one place relation on sentences about individual events, probability is still defined as a property of a class or sequence of sentences, rather than as a property of sentences about just one individual event.<sup>7</sup> Secondly as a one place relation on classes of sentences or a two place relation on classes of events, probability as defined by the standard frequency theory is not relative to evidence - as advocates of the standard frequency theory are fond of pointing out, probability for them is an absolute property of the world independent of whatever 'evidence' might have been gathered.

But, as we have seen, just because the standard frequency theory applies only to classes or sequences, it can not provide an adequate account of single case probabilities and so in the end is to be modified along the lines proposed in Chapter II. There probability was defined as a legitimate, albeit relational, property of individual events. As we will now see, this definition is readily transformed into one whose arguments are pairs of sentences about a given individual event and thus one formally similar to Carnap's definition of probability<sub>1</sub>.

To recast the definition of Chapter II into a form similar to Carnap's definition of probability<sub>1</sub>, we need only replace the pairs of individual events which there were the arguments to our probability function with pairs of sentences stating the occurrence of those events - for this purpose I will adopt the expedient of naming the sentence which states the occurrence of an individual event  $x_n \in A$  etc. by putting single quotation marks around the expression used to designate the event. In Chapter II the probability of an individual event  $x_n \in B$  was made relative to an individual event  $x_n \in A$ ;  $B$  is here the attribute class we are interested in and  $A$  is the reference class required to complete our probability statement. The probability assigned to such pairs of events was determined by the limit of relative frequency  $B$  on  $A$ .

As we have noted before, it has long been recognized by writers on probability that an assignment of probability to an individual event  $x_n \in B$  is equivalent to an assignment of probability to the sentence ' $x_n \in B$ ' asserting



the occurrence of that event. Similarly, of course, the assignment of a relational probability to the pairs of events  $\langle x_n \in A, x_n \in B \rangle$  will be equivalent to an assignment of probability to the pairs of sentences  $\langle 'x_n \in A', 'x_n \in B' \rangle$ . Thus we may offer a definition of probability equivalent to that of Chapter II by defining probability as a function whose domain is a set of ordered pairs of sentences. One sentence of each ordered pair will assert that the experiment in question belongs to the attribute class B we are interested in; the other sentence will state that the experiment belongs to a particular reference class A. The degree of probability assigned to this pair of sentences  $\langle 'x_n \in A', 'x_n \in B' \rangle$  is - in line with the frequency character of the definitions of Chapter II - to be equal to the limit relative frequency of B on A.

The interest in reformulating the definition of Chapter II in this manner is that in this form the relationship between our definition of probability and Carnap's becomes wholly clear. We noted repeatedly that the definition of probability offered in Chapter II made the probability of a particular experiment  $x_n$  having a certain outcome B relative to what reference class  $A_1$ , or  $A_2$  etc., that experiment was regarded as belonging to. In terms of the reformulation just proposed, we may say that the probability assigned to a sentence asserting that an experiment belongs to an attribute class will be relative to sentences asserting what reference class that experiment belongs to. The former sentence  $('x_n \in B')$  is obviously what Carnap would call our 'hypothesis', while the latter sentences  $('x_n \in A_1', 'x_n \in A_2'$  etc.) stating various reference classes to which the experiment belongs obviously are sentences providing different pieces of

evidence for our hypothesis. That is, as we saw in Chapter IV, specifications of what reference class any experiment belongs to are obviously also specifications of evidence about that experiment - evidence to the effect that the experiment fulfills the conditions which define the reference class.

Thus in defining probability as a function whose domain is a 2 place relation on sentences asserting the occurrence of events  $x_n \in A$  and  $x_n \in B$ , we make probability a 2 place relation on sentences that constitute evidence and hypothesis about an experiment  $x_n$ . This is of course just what is done by Carnap's theory of probability<sub>1</sub> as a logical relation of confirmation, and so, as with Carnap, our reformulated definition makes probability a quantitative relation between evidence and hypothesis. Of course as the original definition of Chapter II is an equivalent formulation to the one we are considering now, it also in effect made probability a relation between evidence and hypothesis. Prior to now, however, this fact has been obscured by the original formulation in Chapter II of our definition in terms of a relation between individual events.

In point of fact, the only difference now between our definition of probability and Carnap's definition of probability<sub>1</sub> is that the probability relation obtaining between evidence and hypothesis is, in our theory, measured by empirically given long run frequencies, while for Carnap it is measured by a measure function selected a priori. That is to say the probability, or degree of confirmation, assigned by our theory to the hypothesis ' $x_n \in B$ ' relative to the evidence ' $x_n \in A$ ' is the long run frequency of B on A, the limit of B on A;

for Carnap the probability assigned to such a pair of sentences is determined by an a priori selection of a measure function for the range of these sentences.

Carnap's a priori selection of a measure function is, of course, what makes his theory that of a purely logical relation between evidence and hypothesis. Indeed Carnap makes it perfectly clear that the purely logical nature of the problem he is concerned with rules out selection of a measure function for determining confirmation values that relies on such factual matters as statistical experience of long run frequencies. Citing a suggestion by Waismann that the choice of a measure function be made so as to accord with 'statistical experience' Carnap remarks:

In our theory, on the other hand, the choice of an m-function is regarded as a purely logical question; we shall later define a certain m-function  $m^*$  as the basis for inductive logic. According to our conception, the empirical knowledge of facts enters inductive logic only at one point, viz., as formulated in the evidence  $e$ , but it cannot determine the definition of  $c$  [the confirmation function derived from the m-function]

However, in contrast, on our definition, the degree of probability, or confirmation, an evidence sentence  $e$  provides for a hypothesis  $h$  is very much determined by factual considerations - in particular by long run frequencies, which I have identified with limits of relative frequency.

Here the reader may wish to object that this very identification of long run frequencies with limits in infinite sequences invalidates the claim that the probability values determined on our theory are based on factual



considerations for, of course, we can never know as a fact that a particular sequence has a limit. This objection makes use of the familiar point that limits of relative frequency can never be conclusively verified. However, as I indicated at the outset of Chapter II, I, like many others of a frequentist persuasion, regard the limit of relative frequency in an infinite sequence as a plausible formulation of our ordinary concept of frequency in the long run. Long run frequencies are of course facts of experience and so as a formulation of our ordinary concept of long run frequency, limits of relative frequency are here to be treated as facts of experience. That we can never conclusively verify such 'facts of experience' is at first sight a serious anomaly, but on more careful reflection it appears that unverifiable limits are not wholly different from other so-called 'facts of experience'.

Carnap, for example, in *Testability and Meaning* ably expresses a commonly held view that no sentence about the world can be completely and conclusively verified:

If verification is understood as a complete and definitive establishment of truth then a universal sentence, e.g. a so-called law of physics or biology, can never be verified, a fact which has often been remarked....

Now a little reflection will lead us to the result that there is no fundamental difference between a universal sentence and a particular sentence with regard to verifiability but only a difference in degree. Take for instance the following sentence: "There is a white sheet of paper on this table". In order to ascertain whether this thing is paper, we may make a set of simple observations and then, if there still remains some doubt, we may make some physical and chemical experiments. Here as



well as in the case of the law, we try to examine sentences which we infer from the sentence in question. These inferred sentences are predictions about future observations. The number of predictions which we can derive from the sentence is infinite; and therefore the sentence can never be completely verified. Therefore no complete verification is possible but only a process of gradually increasing confirmation.<sup>9</sup>

Thus the problems which arise in verifying sentences about such ordinary 'facts of experience' as the colour of a piece of paper are distinctly analogous, although not identical, with those of verifying limits of relative frequency. Indeed it is both natural and common to speak of increasingly improved estimates of such limits and so, on analogy to our increasing confidence about simple facts of experience, we may speak of increasing confidence about limits of relative frequency. Now the study of the procedures that yield better and better estimates of limits of relative frequency belongs to statistics proper and the foundation of statistics and, as indicated at the end of the last chapter, these topics are beyond the scope of this thesis; what is important for us is that the unverifiability of such limits is sufficiently analogous to the unverifiability of ordinary facts of experience to allow us to adopt here the common frequentist position that these limits are facts of experience.

One further objection to regarding limits of frequency as facts of experience should be mentioned before we proceed. It is an objection which underlies many discussions of the frequency theory and has recently been formulated succinctly by Mackie<sup>10</sup>. In discussing Von Mises' frequency theory

Mackie remarks:

..... limiting frequencies can never be observed. The statement that tosses of a certain penny have a certain limiting frequency of heads is at best an hypothesis, it has at best the same status as a universal law statement or a law-like statement about an individual object, e.g. that whenever this penny is placed in water it sinks.

This is, of course, essentially the point just discussed concerning the impossibility of verifying a limit statement. However Mackie goes on to explain another sense in which limit statements are hypothetical:

We are not really concerned with any actual sequence of tosses of this penny. Real pennies, unlike the one we supposed above, are never tossed more than, say, a million times. Moreover, they do wear, and it is reasonable to suppose that wear may change our unbiased penny into a biased one or vice versa, i.e. alter its limiting frequency of heads. For these reasons the limiting frequencies we are concerned with are not, as suggested above, hypothesized frequencies in actual sequences, but hypothesized frequencies in hypothetical sequences: the sequence we are concerned with is what we would get if this penny were tossed indefinitely without its present characteristics being changed.

Although it seems to me possible to take such a view of limiting frequency statements, and so develop a frequency theory along these lines, I do not think it is necessary to do so and in general have not done so in this thesis. To do so, as Mackie himself remarks, is to weaken any frequency theory's claim to objectivity for the limits referred to in such theories are not properties of any actual sequence, but rather are properties of, as it were, creatures of our imagination. It seems to me preferable to

take the view that there are certain sequences of events given to us in experience which, although obviously finite in size, can be extended indefinitely.

Any sequence which is extended indefinitely can have a genuine limit of frequency, for a limit of frequency is defined in terms of sequences in which the number of trials  $n$  increases indefinitely, i.e. as  $n \rightarrow \infty$ . Moreover, although we can never be sure that the frequency we have observed so far is the actual limit, it is perfectly possible that the frequency observed so far is equal to the limit in the indefinitely extended sequence.

But what of Mackie's quite correct remarks that no one penny, die etc., is ever tossed indefinitely and so does not form a sequence with an actual limit. The answer here - and it is one that I think is in the back of the mind of most frequentists - is as follows: although any particular penny will not be tossed more than, say, a million times and certainly can not be tossed very much more than this without its present characteristics changing, there are a vast number of other pennies with exactly the same relevant characteristics as this one. Of course no two pennies can have all the same characteristics - if they did they would be one and the same by Leibniz's law - but nor will the same penny have all its characteristics the same on any two tosses. Even if there is no physical change at all between two tosses of the same coin (which is unlikely), each toss occurs at a different time.

What in fact matters in choosing a sequence for determining the probability on an individual toss of a penny is that all the characteristics relevant to the outcome are the same on each toss of the sequence. Over and

above the force with which it is tossed and its initial position - both of which are properties of the toss rather than the penny itself - the only property of a given penny that affects the outcome is its weighting and, of course, this can vary rather widely without any real effect on the outcome. Thus the sequence of repeated tosses of all roughly symmetrical pennies is just as material to this roughly symmetrical penny's probability on a given toss as is the sequence of repeated tosses of it itself.

The sequence of repeated tosses of ordinary roughly symmetrical pennies will of course never contain more than a finite number of elements but, and this is the important point, there is no finite number which necessarily fixes an upper bound to the number of tosses of regular pennies. This is only to say that we can indefinitely extend the sequence of tosses of pennies; I suppose to fully justify this view one would have to argue that the human race will continue indefinitely (not so improbable) and that its members will continue to gamble (again not so improbable) and that inflation will not devalue currency so rapidly that the penny ceases to be minted for commercial use (?), but the general point here should be clear - it is that we are constantly confronted in experience with trials which in all relevant respects are not only repeated frequently but moreover can, if we wish, be repeated indefinitely. Such empirically given and indefinitely extendible sequences may well have a limit of frequency for the outcome we are concerned with and it is in light of our best estimates of this limit that we make judgments of probability. In the ensuing discussion then I will follow over this point the original definition



found in Chapter II and regard the relational probability between sentences ' $x_n \in B$ ' and ' $x_n \in A$ ' as determined by the limit of relative frequency of B in the actual and indefinitely extendible sequence A; for the reasons I have just given above I will continue to regard such limits as 'empirical facts' found in the world as it is.

Since the (empirically determined) relational probabilities between sentences ' $x_n \in A$ ' and ' $x_n \in B$ ' arrived at on our theory can be regarded as relational probabilities for the hypothesis ' $x_n \in B$ ' on the evidence ' $x_n \in A$ ', the difference between our theory and Carnap's can now be expressed concisely by saying that for us probability is a quantitative empirical relation between evidence and hypothesis, while for Carnap it is a quantitative logical relation. Now the proposal to take probability as an empirical, rather than logical, relation between evidence and hypothesis is not original to this writer; indeed, versions of such a proposal have been advocated almost as long ago as the logical relation thesis. No one writer is traditionally associated with such a proposal, but as far back as the early 1930's Waismann<sup>11</sup>, and (in passages) Popper<sup>12</sup>, both quite explicitly argued that probability should be regarded as an empirical relation between evidence and hypothesis, that is, one the numerical values of which are determined by long run frequencies. More recently Kneale<sup>13</sup> has proposed a range theory very similar to Carnap's save that the measures for the ranges he proposes are primarily determined in the light of statistical experience. Yet more recently Mackie<sup>14</sup> has drawn attention to examples in which the probability relations between evidence and hypotheses are empirical.

Now, as we saw, Carnap was fully aware of at least one such proposal to regard probability as an empirical relation between evidence and hypothesis, namely Waismann's suggestion that measure functions for the range of sentences be chosen to accord with our statistical experience of long run frequencies. Why then should Carnap maintain that there are two quite distinct concepts of probability, one based on empirically given long-run frequencies, the other on relations between evidence and hypothesis? Or, perhaps more to the point, why should anyone else accept such a distinction if, as I have tried to show, the long run frequencies found in statistical experience can be regarded as empirical measures of the probability relation between evidence and hypothesis? The answer, of course, is that the concept of probability<sub>2</sub> that Carnap was at pains to distinguish from the concept of a logical relation between evidence and hypothesis was that found in the standard frequency theory. As we noted near the outset of this chapter, the standard frequency theory's concept of a class probability, or limit of relative frequency, is in a number of different respects distinct from that of probability as a relation between evidence and hypothesis. It is a frequency theory for the single case, as developed in Chapter II and reformulated in this chapter, that bears a strong resemblance to Carnap's theory of probability<sub>1</sub>, rather than the standard frequency theory, which indeed is a very different matter.

But if, as I argued in Chapters I - III, the standard frequency theory must be modified because it so egregiously fails to deal properly with the single case, we need no longer accept its very distinct conception

of probability. By abandoning the standard frequency theory in favour of the frequency theory for the single case I have argued for, we also abandon a concept of probability radically distinct from that of a logical relation between evidence and hypothesis for, as we have seen in this chapter, our frequency theory for the single case can be regarded as one which makes probability an empirical relation between evidence and hypothesis.

Now of course no act of philosophical analysis can reduce the concept of limit of relative frequency in an infinite sequence to that of a relation between evidence and hypothesis and so, strictly speaking, one cannot reduce the standard frequency theory's concept of a class probability to that of a relation between evidence and hypothesis. But the real question is whether there is a genuine concept of probability such that probabilities are identical with limits of relative frequency. As acknowledged near the end of Chapter II, we do quite naturally speak of the probability of, say, heads in repeated tosses of a coin (which of course form a class) or, say, the probability of mortality among a certain class of men. Accordingly it is natural to speak of class probabilities; moreover in general we will regard the numerical value of probability in such cases as equal to the frequency in the class in question.

However, this is not to say that these quite legitimate 'class probabilities' are identical with frequencies; rather, as I argued, at the end of Chapter II, it is more plausible to regard these 'class probabilities' as relational probabilities for each of the individual members of the class

in question, where the numerical value of the relational probability is fixed by the frequency in the class. When we spoke, for example, of the probability of mortality  $p$  among 40 year old men we did not mean to ascribe to this class of men some property that no individual man could conceivably have, rather we meant that each person, in so far as he was a man of 40, would have the probability  $p$  of dying within a certain period of time. In terms of the current discussion, we may say that the so-called class probability of mortality among 40 year old men is the probability of mortality for each (and every) person, on the evidence that he is a man of 40. The long run frequency in this class of men is not itself the probability, but is, like Carnap's a priori measure functions, that by which the numerical value of this relational probability is determined.

Put more generally, we found in Chapter II that a class probability of degree  $p$  for an attribute  $B$  in a repeated sequence  $A$  of experiments  $x_1 \dots x_1 \dots$  was not identical with the frequency of  $B$  on  $A$ , but rather was a probability  $p$  numerically equal to this frequency for the pairs of individual events  $\langle x_i \in A, x_i \in B \rangle$  for all  $x_i$  of  $A$ . An obviously equivalent analysis of the class probability  $p$  of  $B$  on  $A$  is that it is a probability of  $p$  for the pairs of sentences  $\langle 'x_i \in A', 'x_i \in B' \rangle$  for every  $x_i$  of  $A$  - here we again simply exploit the equivalence already noted between the assignment of probabilities to events and to sentences asserting the occurrence of those events.



If this analysis of class probabilities is accepted we may readily abandon the view that there is a concept of a statistical 'class probability' distinct from that of probability as a relation between evidence and hypothesis. The quite natural idioms in which a probability for an attribute  $B$  numerically equal to the long run frequency  $p$  in a class  $A$  of experiments is apparently ascribed to  $A$  - statements about so-called class probabilities - are now seen as equivalent to universally quantified statements in which a probability  $p$  is assigned for the hypothesis of  $B$  to each experiment  $x_1$  of  $A$ , on the evidence that  $x_1$  belongs to  $A$ .

In this way Carnap's two distinct concepts of probability are reduced to one more basic one - that of probability as a quantitative relation between evidence and hypothesis about individual experiments. This relation can be measured in two different ways - logically by use of measure functions chosen a priori and empirically by use of the long run frequencies found in classes of repeated events; moreover, of course, this relation is capable of figuring in two kinds of statements - singular, or unquantified, statements as well as universally quantified statements. Statements about the class probability of  $B$  on  $A$  which appeared to justify introduction of a distinct concept of probability are now seen as equivalent to universally quantified statements to the effect that a probability relation measured by the frequency of  $B$  on  $A$  holds between the hypothesis ' $x_1 \in B$ ' and the evidence ' $x_1 \in A$ ' for each  $x_1$  of  $A$ .

Having shown how Carnap's two concepts of probability can thus be reduced to one more fundamental one by abandoning the standard frequency theory in favour of the modification proposed here, we should now consider one objection to the above line of analysis that might be made by an advocate of the standard frequency theory. The objection is that while the frequency definition given here genuinely applies to the single case, it does so at the expense of 'objectivity'. The probabilities assigned to the single case are relational and, in the last (or rather the above) analysis, such relational probabilities turn out to be relative to evidence known about the single case. Thus in contrast to the probabilities of the standard frequency theory which are objective long run frequencies, our probabilities are relative to evidence, which, since evidence is that which is known, depends on the subjective matter of what can be known.

Although, as it stands, I think this objection is misguided, it is simply a variant of an objection that has had some currency among probability theorists. This is the objection that logical relation theories lead to subjectivism. It is somewhat easier to document the existence of this objection than to express it in a convincing manner. Keynes and Jeffreys, who prior to Carnap, offered the most elaborate theories of probability as a logical relation between evidence and hypothesis, seem to be the most common targets for such a charge of subjectivism. Popper simply calls Keynes's theory of probability the 'subjective' theory and the rationale he offers would appear to extend this sobriquet to any who took probability to be a relation between evidence and hypothesis<sup>15</sup>. In a similar vein Von Mises calls

Keynes 'a persistent subjectivist' and, regards the theories of both Jeffreys and Carnap as equally subjective<sup>16</sup>. More recently, Lucas has argued that the logical relation theory, and if I understand him correctly, any theory of confirmation, readily leads to subjectivism<sup>17</sup>.

However, both Keynes and Carnap go to considerable effort to establish the objectivity of their theories. As we noted briefly already in Chapter IV, Carnap's theory is objective in the important sense that the numerical value of confirmation, or probability<sub>1</sub>, given to any pair of evidence and hypothesis is not dependent on actual beliefs about the correct value<sup>18</sup>. Keynes makes the identical point about his theory in the first chapter of his Treatise. A similar, and perhaps stronger, reply can be made to a charge of subjectivism raised against our theory. Probability for us is not a logical relation between evidence and hypothesis, but an empirical one determined by factual considerations of long run frequencies.

Now as I pointed out in Chapter II, this account of probability retains an important feature of the standard theory's account of probability, namely that probabilities are wholly objective entities determined by features of the world as it actually is and so are independent of whatever people might believe about probabilities. On the standard frequency theory these objective probabilities were objective relations between classes of events, viz., limits of relative frequency; on my theory probabilities were relations between individual events but the probability relation which obtains between two

individual events, say  $x_n \in A$  and  $x_n \in B$ , is determined quite objectively by facts in the world, viz., the limit of relative frequency of B on A.

Similarly of course the probability relation between the sentence ' $x_n \in A$ ' as evidence about an experiment and ' $x_n \in B$ ' as hypothesis about that experiment is determined quite objectively by the same fact of the world, the limit of frequency of B on A.

Thus as with the logical relation theory as espoused by Keynes and Carnap, the probability values determined by our theory will not be a matter of any ones' subjective opinion of what this degree of probability is. Furthermore, unlike the probability values determined by the logical relation theory, the values determined by our theory are not given by measure functions selected a priori but by actual facts of experience, viz. long run frequencies. Accordingly, in so far as the charge of subjectivism against logical relation theorists arises from such theories' contention that the degree of probability between evidence and hypothesis is to be determined without regard to other facts of experience, our theory will be found objective in the sense that theirs is not.

As I indicated above, I have found it somewhat easier to locate charges of subjectivism against the logical relation theory than express them clearly and so I can not be sure of the exact grounds for such charges. However, Von Mises and Popper do appear to object to the logical relation theory as subjectivist because its relational probabilities are determined by purely a priori considerations of 'logic' rather than empirically



given frequencies - both writers explicitly state that their objections to the allegedly subjective logical relation theory would be resolved if the degree of probability between evidence and hypothesis were to be determined by long run frequencies rather than the a priori canons of inductive logic.

A good example of this attitude is found in Von Mises' concluding remark on the logical relation theory:

I certainly do not wish to contest the usefulness of logical investigations, but I do not see why one cannot admit to begin with that any numerical statements about probability<sub>1</sub>, about plausibility, degree of confirmation, etc.,<sup>19</sup> are actually statements about relative frequencies.

It should, I think, be clear that our theory of probability as a relation between evidence and hypothesis measured by limits of relative frequency wholly conforms with the conception of Von Mises of what a theory of confirmation should look like; for similar reasons, it coincides with Popper's views as stated in Sections 71 and 72 of The Logic of Scientific Discovery. (Indeed Popper's views in these sections are sufficiently similar to those given here that we find him suggesting, as I have done, that it might be possible to resolve some of the differences between frequency theorists and logical relation theorists by defining probability as a relation between evidence and hypothesis to be measured by long run frequencies).

The charge of subjectivism raised by Lucas against the logical relation theory appears to have a somewhat different basis from that found in Von Mises and Popper and, as far as I can see, nothing said so far in this

chapter suffices to deal with his argument. Lucas bases his argument on the quite correct observation that logical relation theories must construe many ordinary uses of the word 'probable' as elliptical. But, he argues;

..... this will not do. If 'probably' and all related words are elliptical, as alleged, and were short for 'probably - on - the - basis - of - information - at - present - in - my - possession' then there would be no inconsistency at all between my saying 'It will probably rain' and your saying 'It probably will not'. We should be talking about different relations, just as if I were saying 'It is a long way from home' and you were saying 'It is not a long way from home' .... We are in fact talking about the same thing, namely whether it will rain tomorrow or not, and not two different things, namely the relation between the proposition that it will rain on Wednesday and my other beliefs, and the relation between that same proposition and your other beliefs.<sup>20</sup>

Lucas elaborates on this argument and then concludes that for "these reasons logical relation theories readily lead to subjectivism."

Although different from the charge of subjectism already discussed in this chapter, the reader may find a familiar ring to the argument adduced here by Lucas. In point of fact it is closely related to, if not identical with, an observation made in Chapter IV concerning the requirement of total evidence as formulated by Carnap. Carnap, as we saw, construed the evidence available to a person at a given time as that he was aware of at that time; the requirement to take the total available evidence then reduced to the rather weak requirement that one should consider all the information one is aware of. In Chapter IV we noted that this introduced an element of

subjectivism when Carnap's theory was applied to actual situations to determine the unique probabilities required for the purpose of action. The probability value determined for an hypotheses by his relational theory in conjunction with the requirement of total evidence would vary from person to person depending on the subjective matter of just what evidence was known by the person making the probability judgment.

The connection of this remark from Chapter IV to Lucas's argument here is obvious: virtually all logical relation theorists agree that apparently non-relational probability statements are elliptical for one's involving the total evidence available at the time the statement is made. Thus any logical relation theorist who regards the total evidence available to someone when he is making a probability statement as just that of which he is aware, must conclude that apparently non-relational probability statements made by people who know different amounts of evidence will be about different probability relations. This is just the unhappy conclusion Lucas has drawn attention to; moreover, as he appreciates, any theory which makes probability a relation between evidence and hypothesis, whether logical or empirical, can similarly lead to this conclusion.

Now in the first instance it might be thought that the only body of evidence that could be said to be tacitly relied on when someone makes an apparently non-relational probability statement is the totality of evidence known to him. Indeed, the reasoning goes, the only kind of evidence someone could have in mind when he makes an apparently non-relational statement of

probability is, to put it glibly, the evidence in his mind, that is, the evidence actually known to him. It can hardly be supposed, this reasoning continues, that someone making a probability statement believes his statement to be relative to evidence unknown to him. But, as we just saw, such a view of apparently non-relational probability statements leads inevitably to a form of subjectivism.

It is, happily, quite easy to show this reasoning to be mistaken. Consider Lucas's own example of a dispute over the probability of rain for the next day between A and B who have different beliefs about the conditions obtaining at the moment. Lucas seems to me to be correct that A and B are talking about the same thing, and so not talking about the two different relations between the hypothesis of rain and the evidence believed by A versus the evidence believed by B. But what is the same thing they are talking about?

This is readily answered if we consider how A and B might settle their dispute; what in the end they are prepared to agree on is just that thing they were talking about to begin with but, for some reason, disagreed over. A might point out that B, in claiming that rain was improbable, ignored the fact that the air pressure was falling rapidly as shown by a barometer. B might not have known this or worse might have mistakenly believed the pressure to be rising. B, however, while accepting the fact that it is really falling and, to some extent, the implications of this, remains unconvinced, though now slightly less sure of himself. A then points out the presence of a kind of cloud usually found prior to rain. Again perhaps



B failed to notice this. If B still were not convinced, A might, sensibly enough, suggest calling the weather service. They do this and, on hearing the prediction of rain for the morrow, B finally agrees that A was correct all along.

What has A done in persuading B of his mistake? By pointing out the fall in air pressure and the presence of clouds fitting a certain description A has appealed to evidence which can be obtained by implementing extant - and in this case simple - scientific procedures, namely checking a barometer and naked eye observation of cloud formation. This evidence was in a very natural sense available at the time, although as he has to admit, B himself did not know it. Obviously any appeal to information gathered in such a way is germane to their dispute and, in fact, in the end the dispute is settled by an appeal to that institution assumed to know all the evidence that can be obtained by existing relevant scientific procedures, the weather service.

The totality of evidence which can be obtained by implementing all relevant extant procedures is what I called the total evidence available, in the objective sense of 'available' explained in Chapter IV. That A and B are prepared to settle their dispute by an appeal to experts presumably aware of this totality of evidence shows that what in fact they were disputing all along was the probability of rain in relation to the total available evidence, understood in this strong objective sense. The dispute originally arose because unaware themselves of all the evidence in this totality - perhaps

mistakenly believing other evidence - they had different opinions about the exact nature of the probability relation between the hypothesis of rain and this totality of evidence.

Thus contrary to what Lucas claimed, we need not analyze apparently non-relational probability statements made by speakers with different beliefs to be about different relations for the hypothesis in question; rather, as his own example shows, such statements are usually about the same relation, that between the hypothesis and the total evidence available, in our strong objective sense of this word. Thus, Lucas's claim that theories of probability as a relation between evidence and hypothesis lead to subjectivism is now seen to depend on an over-simplification, the simplification that such theories must analyse apparently non-relational statements of probability as elliptical for one's making reference to the totality of evidence known by the person making the statement.

Lucas's argument here illustrates well why many have thought that theories of probability as a relation between evidence and hypothesis lead to subjectivism - and why, on the whole, they were wrong in thinking so. As acknowledged throughout this thesis, we can not simply assert that probability is a relational concept and ignore the fact that we not only often require a unique probability for the purposes of action, but also believe in general there actually is one probability value to be preferred for the purpose of action. We were able to give an account of this fact within the framework of the relational theory put forward in Chapter II by demonstrating that the

probability value for an event relative to the narrowest available reference class would be the most advantageous for the purpose of action. Given the equivalence of probability for events and for sentences asserting the occurrence of those events, as well as the similarity between the principle of choosing the narrowest available reference class and that of using the total available evidence, we can for similar reasons say that the probability for a hypothesis to be preferred for the purpose of action is that relative to the total available evidence.

But if the total evidence available to a person is understood as just that evidence he is aware of, this account of which probability value is to be preferred for the purpose of action entails that different people knowing different amounts of evidence can quite rightly single out different 'preferential' probabilities for the same hypothesis. Since the preferential values decided on for the purposes of action are such an essential part (some would say the essential part) of how and why we make judgments of probability, such an account of the matter leads to a conception of probability that is subjective in a vital respect - the preferential probabilities for the purpose of action this account sanctions do not reflect any objective feature of the world but instead depend on the subjective matter of just how much evidence is known by the person making the judgment of probability.

This I take it is just what, in a back-handed way, Lucas has fastened on to. The probability value preferred for the purpose of action is, of course, the value we speak of in our ordinary and apparently non relational

statements of probability. Thus if apparently non-relational statements of probability are analyzed as relative to the total evidence known to the person making the statement, we are, in effect, regarding the preferential values of probability used for the purpose of action to be those relative to the total evidence known to the person making the probability judgment. As remarked at the end of the above paragraph this does lead to a subjective conception of probability, as Lucas claimed.

However our discussion of Lucas's argument also showed that this charge of subjectivism could only be sustained by taking an oversimplified, and not very plausible view, of the body of evidence which relational theories of probability identified as the one to be used for the purpose of action. If the preferential probability value is identified as the one relative to the total available evidence, when 'available evidence' is defined in the strong objective sense of that which can be known by implementing any relevant extant scientific procedure, the preferential value for a given hypothesis will not vary from person to person depending on what evidence is known. Rather the preferential value for a given hypothesis will be fixed quite objectively for it is the one relative to the totality of evidence which can be obtained by implementing the relevant scientific procedures extant at the time.

This entire discussion of subjectivism can conveniently, and profitably, be brought together by stating quite generally the conditions which need to be fulfilled if a theory of probability as a relation between



evidence and hypothesis is to avoid subjectivism of any kind. First of all, of course, the theory must assign probability values - or as Carnap would call them measure values - to given pairs of evidence and hypothesis in a way that does not depend on mere personal opinion. To have a convenient term we may say that any theory that fulfills this condition is free from subjectivism in the theory proper. As we have seen, both our theory of probability as an empirical relation and the logical relation theory fulfill this condition and so no subjectivism enters into these theories proper.

But as we have just seen at length, a different form of subjectivism can enter when a relational theory of probability is applied to a given situation with the hope of determining unique probability values. Assuming, as most writers do, that the methodological rule we follow in attempting to determine unique probability values is that of using the total available evidence, the results of applying the theory will actually vary from person to person if the total available evidence is understood in the weak sense of the totality of evidence known to the person making the probability judgment. To avoid such subjectively variable results we may still adhere to our principle of using the total available evidence, but we then must define available evidence in a way that is stronger than this sense. The objective criterion of availability elucidated in Chapter IV is, in the first instance, a plausible way of so defining the available evidence. As the reader will have appreciated from the above discussion, treating the evidence to be used in applying, say, Carnap's theory of probability to a

given situation as the total available in this objective sense leads to results that are highly objective. The probability value for a particular hypothesis in relation to the total evidence available in a given situation will be unique and, what is more important for us now, will also be independent of both what evidence happens to be known by particular people and what those particular people believe are the correct probability values to be assigned to given pairs of evidence and hypothesis. The evidence available for a hypothesis will be an objective matter, depending on what scientific procedures are extant, and the number assigned to a hypothesis in relation to the total available evidence will be objectively fixed by the canons of inductive logic.

That we will obtain highly objective results if we apply a theory like Carnap's in relation to the total available evidence understood in this strong objective sense is precisely the point I made in Chapter IV when discussing the doubts raised by Mellor over the objectivity of relational theories of probability. Subsequently, we found that such <sup>an</sup> objective criterion of available evidence was demonstrably not that relied on in all circumstances and in the end distinguished a multiplicity of criteria for what was ordinarily regarded as the available evidence. This suggests that a more intuitively satisfactory way to avoid any subjectivism arising in the application of a relational theory of probability might be to adopt the objective criterion of what is to be regarded as the evidence available in applying the theory to certain circumstances and various, what I called, inter-subjective criteria for the evidence regarded as available in applying the theory to other circumstances. When the amount of evidence

available in a situation is determined by an inter-subjective criterion stronger than the 'subjective' criterion, the application of our theory will arrive at results that are inter-subjectively valid in those circumstances, that is, results that do not depend on just what evidence happens to be believed by particular people; rather they will depend on what can be known, within the bounds of certain practical restrictions, to any number of people.

In any event, by adopting either the objective criterion of available evidence or a mixture of the objective criterion and various inter-subjective criteria depending on circumstances, the principle of using the total available evidence will give unique application in actual situations to theories of probability as a relation between evidence and hypothesis; moreover, the numerical value of these unique probabilities will not be subjectively dependent on what evidence happens to be known by people in that situation. Thus the upshot of this discussion of subjectivism is that it is possible to insure that no element of subjectivism enters into either Carnap's theory of probability as a logical relation between evidence and hypothesis or our theory of probability as an empirical relation between evidence and hypothesis. Both are free from any taint of what I called subjectivism in the theory proper and, by formulating our methodological rule of application along the lines just indicated, we can ensure that the preferential values of probability sanctioned for the purpose of action do not vary from person to person for a given hypothesis depending on what evidence is known about that hypothesis.

Thus, to return to the context in which the topic of subjectivism was first raised, we may be confident that if we abandon the standard frequency theory in favour of the frequency theory for the single case proposed here no new element of subjectivism will enter probability theory as the standard frequency theory exits. Accordingly - and this brings us full circle back to the original topic of this chapter - we will be fully justified in abandoning the standard frequency theory in favour of one in which probability is treated as an empirical relation between evidence and hypothesis - a relation measured by long run frequencies - and in doing so we find ourselves with one fundamental concept of probability as a relation between evidence and hypothesis in place of the two distinct concepts originally articulated by Carnap.



## Footnotes to Chapter V

1. Hacking, I., The Emergence of Probability, Cambridge 1975, Chap. 2.
2. Mackie, J. L., Truth, Probability and Paradox, Oxford 1973, Chapter 5.
3. Hacking, I., LSI, p.15.
4. Gillies, D., op. cit. p.31.
5. Ayers, M., The Refutation of Determinism, London 1968, pp. 46-50.
6. Mellor, H.D., op. cit., p.1.
7. Carnap, R., LFP, Section 10B.
8. Ibid, p.299.
9. Carnap, R., Testability and Meaning, Readings in The Philosophy of Science, ed. Feigl and Brodbeck, New York 1953, p.48 ff.
10. Mackie, J. L., op. cit. p.177.
11. Waismann, F., Logische Analyse des Wahrscheinlichkeitsbegriffs, Erkenntnis, I (1930-1931) pp. 228-248.
12. Popper, K., op. cit., Sections 71 and 72.
13. Kneale, W., Probability and Induction, Oxford 1949 Sections 35-37.
14. Mackie, J. L., op. cit., p.164.
15. Popper, K., op. cit., pp. 148-9.
16. Von Mises, R., op. cit., pp. 94-96.
17. Lucas, J.R., The Concept of Probability, Oxford 1970, pp. 52-56.
18. Carnap, R., LF, Section 72.
19. Von Mises, R., p.97.
20. Lucas, J.R., op. cit., pp. 53-54.

## Chapter VI

### The Principle of Indifference and The Classical Theory of Probability

Having thus shown that the concept of a relation between evidence and hypothesis underlies the statistical and epistemic concepts of probability originally distinguished by Carnap, I will now examine the classical theory of probability and the principle of indifference to see if, taken together, they constitute a concept of probability distinct from this basic one so far discussed. In fact I will argue in this chapter that the classical theory taken together with the principle of indifference should be understood as the first and therefore, not surprisingly, somewhat muddled attempt to put forward a version of just this definition of probability as a relation between evidence and hypothesis. Of course many writers have pointed out affinities between the classical theory of probability and the modern logical relation theory. Carnap, for example, cites numerous passages from classical writers to support his general contention that the explicandum they had in mind is his concept of probability<sub>1</sub>.<sup>1</sup> But the position I wish to maintain is considerably stronger than this and it depends on a detailed analysis of the controversial principle<sub>2</sub> of indifference. What I wish to claim is that the principle of indifference is itself really an early and rudimentary version of the definition of probability as a relation

between evidence and hypothesis, a version in which probability is defined as the comparative concept of evidential support. Moreover, I hope to be able to show that most of the well-known objections to the principle are natural, but not insoluble objections, to the definition of probability as a relation between evidence and hypothesis incapsulated in the principle and to the rudimentary and, to some extent, muddled way this definition is expressed by the principle. Indeed the analysis of the principle given in this chapter - and the lengthy discussion of objections to the principle of the next chapter - is offered primarily in the hopes of illuminating and resolving some of these long-standing objections to it.

At first sight it might appear strange to construe the principle of indifference as in any sense an instance of the definition that probability is the relation of evidential support. Certainly its usual formulation to the effect that alternatives are equally likely if and only if there is no reason to choose between them does not, on the face of it, look like a definition of probability. To see the sense in which the principle of indifference should be understood as a definition of probability akin to the definition of probability as the relation of evidential support I have already discussed, we need to refer to an important insight of Frege on the nature of semantic definitions, namely the insight that a perfectly proper - perhaps the only proper - way to provide the semantics for a concept is to state identity conditions for that concept. Thus, I will claim, in just the way the statement of the conditions under which

classes\* were equal in number constituted the essence of Frege's definition of number, the principle of indifference can be understood as offering a semantic definition of probability in so far as it fixed the conditions under which different alternatives were equal in probability.

Let me hasten to add that I do not wish to maintain that previous proponents of the principle of indifference have put it forward as a definition of probability in the sense that I believe it is; indeed if such an interpretation of the principle already existed there would be little point in my arguing for it here. Rather my contention is that both proponents and opponents of the principle have failed to see its essentially definitional character and in consequence have to a large degree failed to appreciate what are the real issues involved in the controversies surrounding the principle. Certainly such a situation would not be without precedent in the history of science; until the definitional and empirical components of Newton's Third Law of Motion were clearly separated and identified by Mach, one could not properly evaluate various objections to Newton's introduction of the concepts of 'force' and 'mass' in his theory. Once the definitional and empirical components of the Third Law were separated it became clear that critics of Newton were correct if they maintained that inertial mass (and thus by the Second Law force as well) was not a concept that could be

---

\*Or rather for him 'concepts' in his extensional sense of that expression.



given a meaning independently of the Third Law, but were incorrect in denying him the right to introduce a concept that could be operationally defined and which, moreover, occurred in a set of laws that had various testable consequences.

In a somewhat similar vein much of this and the subsequent chapter will be aimed at clarifying the issues in question in previous discussions of the principle of indifference by arguing for the definitional status of the principle; accordingly, we should begin with at least a quick look at some of the variety of views put forward concerning the principle. To begin with we find writers who claim it to be analytically valid in virtue of what the term 'probability' means:

If there is no reason to believe one hypothesis rather than another, the probabilities are equal. In terms of our fundamental notions of the nature of inductive inference, to say that the probabilities are equal is a precise way of saying that we have no ground for choosing between the alternatives. All hypotheses that are sufficiently definitely stated to give any difference between the probabilities of their consequences will be compared with the data by the principle of inverse probability; but if we do not take the prior probabilities equal we are expressing confidence in one rather than another before the data are available, and this must be done only from definite reason. To take the prior probabilities different in the absence of observational reason for doing so would be an expression of sheer prejudice. The rule that we should take them equal is not a statement of any belief about the actual composition of the world, nor is it an inference from previous experience; it is merely a formal way of expressing ignorance . . . . . It is not a new rule in the present theory because it is an immediate application of Convention 1 [that 'We assign the larger number on given data to the more probable proposition (and therefore equal numbers to equally probable propositions)]<sup>2</sup>.

The reader may be excused for failing to see exactly how the principle of indifference follows directly from Jeffrey's Convention 1 as the convention presupposes that we have already determined certain propositions to be equally probable - and that is exactly what the principle of indifference is intended to provide a criterion for. Be that as it may, it is clear that Jeffreys sees the principle as following directly from the meaning of the term 'probable' and there is, to my mind, little doubt that there is some intuitive plausibility to this claim. The intuition that the principle borders on the analytic is felt most strongly if we try to think of a counter-example to it, that is, two propositions that we hold to be equally probable, despite having reason to prefer one to the other. The difficulty of finding such a counter-example readily convinces us that, as Keynes puts it,

... the principle certainly remains as a negative criterion; two propositions cannot be equally probable, so long as there is any grounds for discriminating between them.

On the other hand, there are those, particularly of a frequentist persuasion, who, far from accepting the principle as analytic, whole-heartedly reject it - and not without a certain apparent justification. Von Mises, for example, cites the famous argument that by using the principle of indifference one can arrive at the seemingly absurd result that any proposition of which we know nothing other than that it or its negation may be true is  $\frac{1}{2}$  probable, because by hypothesis, we have no reason to choose between the two alternative propositions. He then goes on, in reply to Keynes's attempt to resolve this and related difficulties to argue:

It does not occur to him [Keynes] to draw the simple conclusion that if we know nothing about a thing, we cannot say anything about its probability.<sup>4</sup>

Between the extremes of outright acceptance and outright rejection, we find numerous writers who, while admitting that an unrestricted application of the principle can lead to absurdities and contradictions, think it has sufficient merit for some substantive reformulation of it to render it acceptable. The details of the reformulations proposed vary considerably; for the moment we will restrict ourselves to those of Kneale, Keynes and Carnap.

Beginning rather in the vein of the remark 'Ex nihilo nihil' made as early as 1842 by Ellis<sup>5</sup>, Kneale says, 'Probability statements may be modest assertions, but even they can not be justified by mere ignorance'<sup>6</sup>. Kneale then attempts to reformulate the principle so as to base judgements of equal probability on the knowledge of absence rather than the absence of knowledge:

I have argued that we are entitled to treat alternatives as equiprobable if, but only if, we know that the available evidence does not provide a reason for preferring any one to any other. According to the principle of indifference we may call alternatives equiprobable if we do not know that the available evidence provides a reason for preferring any one to any other. Instead of knowledge of absence Laplace and those who agree with him accept absence of knowledge as a sufficient ground for judgments of probability. This change accords with their subjectivism. For 'absence of knowledge' signifies only a fact about the mind, whereas 'knowledge of absence' signifies not only a fact about a mind, but also a truth about something independent of that mind. The same point can be put in another way by distinguishing two senses of the word 'indifferent'. According to the principle of indifference alternatives are equally probable if I am indifferent in my attitudes towards them. According to the theory I have put forward it is necessary that the alternatives themselves should be indifferent i.e. without difference in a certain respect.<sup>7</sup>



The charge of subjectivism leveled by Kneale at the principle is, of course, a common complaint, put forward by many others. For the moment the merits of this objection need not detain us - we note it only for the purpose of conveying something of the range of not implausible objections to the principle. Within this range we find writers, who while raising other objections to the principle, do not find its alleged subjectivism problematic. Keynes, for example, acknowledges that unless certain precautions are taken, the principle leads to outright contradictions; however, once the principle's application has been restricted in such a way as to avoid these contradictions, Keynes thinks the principle wholly acceptable. He begins his famous chapter<sup>8</sup> on the principle by saying:

The rule as it stands may lead to paradoxical and even contradictory conclusions. I propose to criticise it in detail and then discover whether any valid modification of it is discoverable.

A particularly simple version of the contradictions to which Keynes refers is his now classic example of the probability of the colour of a given book.

Thus [by the principle of indifference] if  $a$  and  $\bar{a}$  are contradictories, about the subject of which we have no outside knowledge, it is inferred the probability of each is  $\frac{1}{2}$ . In the same way the probabilities of two other proposition,  $b$  and  $c$ , having the same subject as  $a$  may be each  $\frac{1}{2}$  ... If, for instance, having no evidence relevant to the colour, we could conclude that  $\frac{1}{2}$  is the probability of 'This book is red', we could conclude equally that the probability of each of the propositions 'This book is black' and 'This book is blue' is also  $\frac{1}{2}$ . So that we are faced with the impossible case of three exclusive alternatives all as likely as not. [This quotation, like most others from Keynes, is to be found in Chapter IV of the Treatise; unless otherwise stated all references to Keynes in the remaining parts of this thesis are from this chapter.]



After leading us through a maze of similar contradictions, engendered in large part, as we shall see later, by dividing up a given problem into different sets of mutually exclusive and exhaustive alternatives, Keynes proposes, for non-geometrical probability, that the principle be only applied to those cases where the alternatives are 'ultimate'. For geometrical probabilities, his restrictions for avoiding difficulties are more complicated but, he assures us, if 'we are careful to enunciate the alternatives in a form to which the principle can be applied unambiguously we shall be ... able to reach conclusions in geometrical probability which are unambiguously valid.'

We should note that Keynes never explicitly characterizes the logical status of his resurrected principle of indifference. Despite the existence in Part II of the Treatise of an elaborate classificatory system of definitions, axioms and theorems for the fundamental principles of the probability theory, the principle evades classification. Unlike Jeffreys, who justifies it twice by stating it to be at once a direct application of a fundamental convention and also an analytic consequence of the terms involved, Keynes only characterizes it as 'the most important' rule 'by means of which the probabilities of different arguments can be compared.'<sup>9</sup>

Just such status as a rule of somewhat unclear logical lineage seems to have been the fate of the principle of indifference at the hands of most of its proponents. Only with the advent of the heightened standards of rigour imposed on the philosophy of science by advances in syntax and semantics have we seen attempts made to classify the logical status of the

**Principle of Indifference.** Carnap's pioneering work on Inductive Logic is perhaps the best example of a system in which the logical status of the principle becomes clear.

Carnap believed that what in most previous writers appeared to be a rule for determining probabilities would be in his more rigorous system a measure function  $m$  (leading directly to a confirmation function  $c$ ) on the atomic sentences of the language in question. Although this does, of course, seem the most natural status to ascribe to the role traditionally played by the principle of indifference, it is not the status that I think correctly characterizes it, nor, in fact, as I will argue later, is it one that best fits with certain rather acute remarks that Keynes eventually makes concerning the principle.

But first let us look in some detail at Carnap's treatment of the principle for it provides us with at least one perfectly clear statement of the logical status of the principle of indifference construed solely as rule for determining probability values. Carnap in his Logical Foundations of Probability and The Continuum of Inductive Methods intends to explicate what he calls the concept of probability<sub>1</sub>, probability as the logical relation of evidential support. The degree of confirmation -  $c(h, e)$ , or probability<sub>1</sub>, assigned to a hypothesis  $h$  on the basis of evidence  $e$  is determined by the ratio of the measure of the evidence and the hypothesis to the measure of the evidence alone i.e.

$$c(h, e) = \frac{m(e, h)}{m(e)}.$$

Thus to fix a confirmation function and thereby explicate the concept of probability<sub>1</sub> we need only determine a measure function for the

sentences of our language. In keeping with his range conception of probability, Carnap stipulates that the measure  $m(j)$  for a sentence  $j$  is the measure assigned to the range of  $j$ . The range of a sentence  $j$  is the set of state descriptions in which  $j$  holds, with a state description defined as an assignment of one from each set of exclusive and exhaustive primitive predicates (and their negations) to each individual constant of the language. Thus with a language with one primitive Predicate  $P$  and two individual constants  $a$  and  $b$ , there are four state descriptions:  $Pa$ ,  $Pb$ ,  $\sim Pa$ ,  $\sim Pb$ , and  $\sim Pa$ ,  $\sim Pb$ . As the measure assigned to a range, i.e. set of state descriptions, is the sum of the measures of all the constituent state descriptions, the determination of a measure for a sentence  $j$  depends on determining a measure for state descriptions. As we obtain a confirmation function directly from measures on atomic sentences, the problem of finding a confirmation function for  $h$  and  $e$  is eventually reduced to the question of finding a measure function for the individual state descriptions of the language.

Enter now the principle of indifference. Carnap cites one of the long standing controversies surrounding the principle of indifference: whether individual distribution (constitutions) or statistical distributions (ratios) are to be assigned equal probabilities in the absence of relevant information. To understand this controversy and Carnap's treatment of it, think of an urn of which we know nothing except that it contains two balls, each white or not-white. By the principle of indifference we might argue that in the absence of any reason for preference, ball  $a$  is as likely to be

white as not-white and the same for ball b. Reading W for white and  $\sim W$  for not-white, this renders us four distinct possibilities each equally likely:

| <u>Ball a</u> | <u>Ball b</u> |
|---------------|---------------|
| W             | W             |
| W             | $\sim W$      |
| $\sim W$      | W             |
| $\sim W$      | $\sim W$      |

Here we hold each individual distribution or constitution equally likely (note for future reference, the similarity of individual distributions to state descriptions). Alternatively in the absence of any relevant information, we might reason that the over-all statistical distributions (ratios) of two not-whites, one not-white and one white, and two whites are each equally likely. This contradicts our previous result in that the combination of two white balls would now have a probability of 1 in 3, rather than 1 in 4. Carnap rightly remarks that as the debate stands neither side is correct, for either the assumption that individual distributions are equally likely or that statistical distributions are equally likely can be shown to be internally inconsistent unless substantially qualified.

To see this inconsistency consider what happens if we find out that there are exactly two distinct ways a ball of the urn can be not white, either by being black or red. By assuming individual distributions to be equally likely, we would then conclude that there are nine distinct and equally likely possibilities:



| Ball a | Ball b |
|--------|--------|
| B      | B      |
| B      | W      |
| B      | R      |
| W      | W      |
| W      | B      |
| W      | R      |
| R      | B      |
| R      | W      |
| R      | R      |

Here by holding individual distributions to be equally likely we reach the conclusion that there is a 1 in 9 chance of both balls being white, whereas by holding individual distributions equally likely and considering only the two predicates 'white' and 'not-white' we arrived at a figure of 1 in 4 for this result. Obviously the probability of there being two white balls varies considerably depending on which set of predicates are used for the basis of determining the equally likely individual distributions. This line of argument, which is simply one instance of the contradictions to which the principle of indifference can give rise, can easily be applied to show that, with our three colours, each of the 6 statistical distribution (e.g. 2 reds, 1 red 1 black ... etc.,) is equally likely.

Carnap concludes that such difficulties can only be met by using consistently one set of exhaustive and mutually exclusive predicates to describe a given situation. Not unnaturally he favours using those sets of exclusive and exhaustive predicates that explicitly involves all the relevant primitive predicates of the language - the set of 'W', 'R', and 'B' in our example. In Carnap's system the matter is expressed in terms of Q divisions;

unfortunately the definition 'Q divisions' for languages with any finite number of primitive predicates is fairly complicated and it would be inappropriate to explain here all the details of this definition. However in a language with one set of exclusive and exhaustive predicates, it is quite easy to see what a Q division is and so it will be convenient throughout this discussion to keep in mind our language with 'W', 'R' and 'B' as the primitive predicates. A division is defined by Carnap as a set of predicates (not necessarily primitive) such that every individual constant of the language has one but only one of the predicates of that set. In our language with primitive predicates 'W', 'R', and 'B', the set of (molecular) predicates  $\{M_1, M_2\}$  where  $M_1 = W$  and  $M_2 = R \vee B$ , form a division, and these two predicates of course are the troublesome predicates 'white' and 'not-white' from above. But the set of predicates  $\{Q_1, Q_2, Q_3\}$  where  $Q_1 = W$ ,  $Q_2 = R$ ,  $Q_3 = B$  also forms a division and since this division involves explicitly all the primitive predicates of the language, it is what Carnap calls a Q-division. Carnap diagnoses the problem we have just considered as arising from an application of the principle of indifference to different divisions, one formed by the predicates  $M_1$  and  $M_2$ , the other formed by  $Q_1$ ,  $Q_2$  and  $Q_3$ . He stipulates that the problem is to be resolved by applying the principle only to Q-divisions formed by the predicates  $Q_1$ ,  $Q_2$ , and  $Q_3$ .

But this still leaves us with the question of whether, by the principle of indifference, we should assign equal probabilities to individual distributions under the Q-division or to statistical distributions under the Q-division. In Carnap's system individual distributions under the Q-division

constitute state descriptions. Similarly statistical distributions under the Q division correspond to Carnap's 'structure descriptions'. Thus, in order to avoid contradiction by using different sets of predicates (different 'divisions'), the traditional debate surrounding the principle of indifference in regard to individual distributions and statistical distributions reduces for Carnap to the question of whether state descriptions or structure descriptions should be given an equal weight. Thus we find:

The controversy [surrounding the principle of indifference] concerned the question as to which of the following rules should be accepted:

- (A) Individual distributions have equal  $m$  [measure] values
- (B) Statistical distributions have equal  $m$  [measure] values.

Now it can easily be shown that either rule leads to contradiction if taken in the given unrestricted form and hence applied to all divisions.

However, each of the two rules becomes consistent if restricted to any one division. It seems natural to take here for each system the strongest division possible in the system, which is that formed by the Q's. The individual distributions for all the individuals of the system with respect to the Q's are the state descriptions. The corresponding statistical descriptions ... [are] structure descriptions ... Thus the two modified rules are as follows:

- (A') State descriptions have equal  $m$  values [measures]
- (B') Structure descriptions have equal  $m$  values [measures] <sup>10</sup>

The next point to note is that rules (A') and (B') are not, for Carnap, merely cleaned up versions of the principle of indifference, they are also the most natural rules for a measure function used to explicate the concept of probability<sub>1</sub>. That is, as we noted earlier Carnap required only a measure function that determined values for the state descriptions in which the hypothesis  $h$  and the evidence  $e$  held in order to construct his confirmation function for determining the probability of  $h$  on  $e$ . Rules (A') and (B') derived from the principle of indifference are Carnap's first proposals for such a measure function. (A') obviously determines measures for each state description by holding all state descriptions equal and taking, as one must, the sum of the measure of all state descriptions to be 1. (B') also determines measures for all state descriptions, but in a slightly less direct manner - each statistical distribution or structure description is assigned an equal measure e.g. in our example of balls that can be red black or white only the measure of  $\frac{1}{6}$  would be given to each structure description '2 white balls', '1 white ball 1 red', '2 black balls' etc. To determine measures for the state descriptions of this example we must, according to Carnap, apportion out the measures for each structure description equally among the state descriptions that could lead to that structure description (To use Carnap's terminology, we divide the measures of the structure descriptions by the number of isomorphic state descriptions). That is to say, the state description: 'Wa. Wb' receives a measure of  $\frac{1}{6}$  as that is the only state description by which the structure description 2 white balls can be achieved. The state description 'Wa. Rb' is given a measure of  $\frac{1}{12}$  as it is one of two state descriptions that leads to the structure description of 1 white 1 red ball the other being 'Ra. Wb.'



The important point to note, in understanding Carnap's enterprise, is that the probability of drawing a second white ball, after a first one, will be very different depending on whether we base our measure function (and thus confirmation function) on  $(A')$  rather than  $(B')$ . Taking  $h$  as 'Wb' and  $e$  as 'Wa' we see that a measure function  $m^+$  based on  $A'$  yields  $m^+(e) = \frac{1}{3}$

( $e$  holds in three of nine state descriptions, each given an equal measure by  $(A')$ ); similarly  $m^+(e.h) = \frac{1}{9}$ . Thus, as the confirmation  $c^+$ , follows directly from the measure function  $m^+$  by  $c^+(h,e) = \frac{m^+(e.h)}{m^+(e)}$ , we have  $c^+(h,e) = \frac{1}{3}$ .

But using rule  $(B')$  to establish a measure function  $m^*$ , we see that  $m^*(e) = \frac{1}{6} + \frac{1}{12} + \frac{1}{12} = \frac{1}{3}$  as  $e$  holds in three state descriptions,

'Wa. Wb' and 'Wa. Bb', and 'Wa. Rb' given, respectively, measures of  $\frac{1}{6}$ ,  $\frac{1}{12}$ ,  $\frac{1}{12}$  by rule  $(B')$ . As  $m^*(e.h) = \frac{1}{6}$ , and  $\frac{m^*(e.h)}{m^*(e)} = c^*(h,e)$ , we

have  $c^*(h,e) = \frac{1}{2}$ , using rule  $(B')$  for our measure function. In general

rule  $(A')$  leads to assignments of probability wherein the probability of a hypothesis  $h$  on evidence  $e$  is independent of the content of the evidence  $e$  - the reader may confirm this by applying rule  $(A')$  to the proposition  $e'$ : 'Ba' and finding the value of  $c^+(h,e') = \frac{1}{3}$  again. However, the confirmation function  $c^*$  based on  $(B')$  leads to probability assignments wherein the probability of a hypothesis  $h$  concerning the occurrence of a property increases in virtue of evidence of previous instances of that property.

This fact allows Carnap to adjudicate between his two improved versions of the principle of indifference, rules (A') and (B'). Carnap notes that many prominent writers 'have accepted the principle of indifference in the form of (A'), among them Keynes and Wittgenstein'.<sup>11</sup> He also notes that the derived confirmation function  $c^+$  is 'simple, natural' and 'at first glance quite plausible'.<sup>12</sup> However as the  $c^+$  determines probability values for hypothesis on evidence irrespective of the content of the evidence

... the choice of  $c^+$  as the degree of confirmation would be tantamount to the principle never to let our past experience influence our expectations for the future. This would obviously be in striking contradiction to the basic principle of all inductive reasoning.<sup>13</sup>

But using the principle of indifference in form (B') leads to the measure function  $m^*$  and the confirmation function  $c^*$ ; the confirmation function  $c^*$  has just the properties desired for inductive reasoning, and so the  $m^*$  measure function is, for Carnap, the acceptable form of the principle of indifference:

The preceding considerations show that the following argument, admittedly not a strong one, can be offered in favour of  $m^*$ . Of the two  $m$  functions which are most simple and suggest themselves as the most natural ones [via the principle of indifference],  $m^*$  is the only one which is not entirely inadequate.<sup>14</sup>

The essential point here is that Carnap sees the principle of indifference as the basis for measure and confirmation functions used to explicate his concept of probability<sub>1</sub>. For him one fairly standard interpretation of the principle yields an apparently plausible but essentially unsatisfactory

measure of confirmation function, while another well-known interpretation does lead to an adequate measure function and confirmation function.

It may be thought, and I suspect most people do think, that Carnap's  $c^*$  function for inductive logic owes more to Carnap's ingenuity than to the principle of indifference. However, not only is it clear that Carnap himself saw the  $m^*$  and  $c^*$  functions as derived from the principle of indifference, it is also clear that prior to Carnap it had been recognized that probability assignments based on giving equal weights to statistical distributions would lead to results consonant with inductive reasoning - Keynes cites Poisson as offering just such an analysis.

Lest the reader become so convinced at this point by Carnap's interpretation of the principle of indifference that he has little curiosity in reading further, let me hasten to point out that Carnap's proposal, rather than vindicating the assignment of equal measures to structure (or statistical) descriptions at the expense of equal measures to state (or individual) descriptions, actually tacitly relies on the assumption, at certain crucial points, that state descriptions are to be given equal measures. That is, with Carnap's  $m^*$  function each state description that is an instance of a given structure description is assigned an equal measure - the measure for each state description is the measure for the structure description of which it is an instance, divided by the number of isomorphic state descriptions. But once we have abandoned the idea that we can use the principle of indifference to justify assigning equal measures to all

state descriptions, it becomes difficult to see the justice in assigning equal measures to all state descriptions that fall under a given structure description.

However, we need not for the moment consider further the merits of Carnap's derivation of the  $m^*$  function from the principle of indifference. We need only note, at the pain of some repetition, that Carnap quite clearly interprets the principle of indifference - both for his own work and in the writings of others - as an (admittedly somewhat ambiguous) proposal for a measure function  $m$  determining a confirmation function  $c$ , which function  $c$  then may serve to explicate his concept of probability<sub>1</sub>.

Having now surveyed some of the historically significant treatments of the principle of indifference in a spectrum which extends from outright rejection to acceptance as, somehow, analytically valid, I now wish to offer the alternative analysis indicated earlier. To repeat what I said there, my basic position is that the principle of indifference is to be understood as the first, and therefore crude and somewhat muddled, attempt to define probability as the relation of evidential support between evidence and hypothesis. This view is not one that I know to have been explicitly advocated before and, indeed, in large measure it conflicts with the kinds of views towards the principle that we have just canvassed (with the possible exception of Jeffreys' rather sketchy remarks concerning the analyticity of the principle).

To appreciate the definitional aspect of the principle it is necessary to realize that we may never speak of there being 'no reason to prefer' one alternative to another simpliciter, but rather it is only in relation



to some body of evidence that we speak of there being no reason to choose between them. This point is accepted, either explicitly or implicitly, by numerous writers on the principle. Keynes, for example, begins his discussion of the principle thus:

The Principle of Indifference asserts that if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability.

Full justification for the view that only in relation to some body of evidence may we speak of reason to prefer one hypothesis to another can be found if we inquire carefully into the meaning of our intuitive concept of 'reason'. It seems clear that in regard to epistemic matters - which alone concern us here - it is only some body of relevant evidence that can provide or constitute a reason for a particular hypothesis. More exactly we may say that in ordinary usage it is only some body of evidence which supports a hypothesis more than its negation that is called a reason in favour of accepting that hypothesis. Conversely, evidence which supports the negation of a hypothesis more than the hypothesis itself constitutes or provides a reason against that hypothesis.

This characterization of 'reason for' a hypothesis in terms of extent evidence supporting that hypothesis can readily be extended to explain the crucial idiom of having reason to prefer one hypothesis to another. Since it is some extant body of supporting evidence that alone can constitute a reason in favour of a hypothesis, only some body of evidence which supports a

hypothesis  $q_1$  more than  $q_2$  can constitute a reason for preferring  $q_1$  to  $q_2$ . This can easily be generalised to the sets of mutually exclusive and exhaustive alternative hypotheses  $q_1 \dots q_n$  that provide the primary application for the principle of indifference: a reason for preferring  $q_1$  to any of  $q_2 \dots q_n$  is, quite simply, some body of evidence which supports  $q_1$  more than it supports any of  $q_2 \dots q_n$ .

Moreover it should be clear on reflection that while any evidence which supports a hypothesis  $q_1$  more than each of  $q_2 \dots q_n$  constitutes or provides a reason for preferring  $q_1$  to each of  $q_2 \dots q_n$ , we would not ordinarily say we have reason to prefer  $q_1$  to each of these unless the total available relevant evidence, taken as a whole, supported  $q_1$  more than each of them. That is to say, one would obviously be arguing in a fallacious manner if, in the face of overwhelming evidence in favour of a hypothesis  $q_1$ , it was maintained that there was reason to prefer, say,  $q_2$  because some small part of the over-all evidence did in fact support  $q_2$  more than  $q_1$ . It is the evidential support provided by the total available relevant evidence taken as a whole, rather than isolated parts of it, that constitutes the grounds of reason for preference among competing hypotheses.

(That it is the total available evidence taken as a whole that constitutes reason for preference among competing hypotheses seems to me a crucial element in giving a correct analyses of the principle of indifference and one that I will discuss at length later in this chapter. Now as we saw in

previous chapters there are a number of different senses of the expression 'available evidence', ranging from a subjective one in which 'available evidence' means known evidence to an objective one in which the evidence available is fixed by what procedures are extant in science. Given these different senses of 'available evidence' there is admittedly a good deal of vagueness in my assertion that it is the total available evidence taken as a whole that constitutes grounds for preference, for it is not yet clear in which of these senses I here understand available evidence. As undesirable as it may at first seem, I wish to leave the matter vague for sometime further and then discuss the matter later in full. Thus for the time being I will continue to say, somewhat vaguely, that it is the evidential support provided by the total available evidence, taken as a whole, that constitutes the grounds of reason for preference among competing hypotheses.)

Characterized in this way it then becomes apparent that the idiom of 'reason to prefer' which figures so prominently in the principle of indifference is nothing other than an intuitive expression of what Carnap has called the comparative concept of confirmation or, as we may alternatively call it, the comparative concept of evidential support - with it understood that the evidence occurring in these comparisons is the total available. The concept of evidential support at issue is, I think, clearly comparative in that when we judge there to be reason for preference among hypotheses we simply are judging the available evidence to give unequal degrees of support to those



hypotheses. The basic point at issue here, which is in my opinion essential to unravelling the controversies surrounding the principle, is that the truth conditions for the idiom of having 'reason to prefer' one hypothesis to another are those in which we judge the available evidence to provide a greater degree of support to that hypothesis than to the other. Similarly when we speak of there being no reason to prefer one hypothesis to another, we only mean that what evidence there is supports each hypothesis equally well. This analysis applies as well to the problematic case - to be discussed in detail later - wherein the available evidence provides no information relevant to the hypotheses in question. In such cases the ascription of equal probability values to the hypotheses by the principle is again based on judgements of equal, and here null, evidential support.

In this light, it should be clear that the principle of indifference merely states that alternative hypotheses are equally likely if and only if the hypotheses are given an equal degree of evidential support by the extent, or total available, evidence. To put it more generally, the principle of indifference simply specifies the identity conditions for judgments of probability in terms of comparative judgments of evidential support where the evidence used in the comparisons is the total amount available. From this interpretation of the principle, it is a short step - and one we will take shortly - to the conclusion that the principle of indifference is primarily definitional in character for, as clearly seen by Frege, it is the essential feature of a definition that it provides a criterion of identity for the concept to be defined.



First, however, let us note that, interestingly enough, the view propounded here of the principle of indifference as a statement which fixes identity conditions for probability judgments in terms of comparative judgments of evidential support can be extracted almost intact from Keynes' trenchant analysis of a 'reason for preference.' Beginning a line of argument that includes, near its end, the passage from the Treatise most recently cited,

Keynes says:

The principle states that there must be no known reason for preferring one set of alternatives to any other. What does this mean? What are reasons, and how are we to know whether they do or do not justify us in preferring one alternative to another? I do not know any discussion of Probability in which this question has been so much as asked. If, for example, we are considering the probability of drawing a black ball from an urn containing balls which are black and white, we assume that the difference of colour between the balls is not reason for preferring either alternative. But how do we know this, unless by a judgement that, on the evidence in hand, our knowledge of the colour of the balls is irrelevant to the probability in question. We know of some respects in which the alternatives differ, but we judge that a knowledge of these differences is not relevant ... Before, then, we can begin to apply the Principle of Indifference, we must have a number of direct judgements to the effect that the probabilities under consideration are unaffected by inclusion in the evidence of certain particular details. We have no right to say of any known difference between the two alternatives that it is no reason for preferring one of them, unless we have judged that a knowledge of this difference is irrelevant to the probability in question.

Keynes then goes on to define more exactly what is meant by 'irrelevance' and, by contrast, relevance. Although the theory of relevance can be quite complex we need only consider here Keynes's 'simplest definition of irrelevance,' which still forms the basis of the theory of relevance:

$h_1$  is irrelevant to  $x$  on evidence  $h$ , if  
the probability of  $x$  on  $h h_1$  is the same  
as its probability on evidence  $h$

Evidence  $h_1$  is relevant to  $x$  if and only if it is not irrelevant. We should note that in this account, as in others, irrelevance and, thus relevance, is defined in terms of judgments of equal probability viz.,  $h_1$  is irrelevant to  $x$  on  $h$  if and only if the probability  $x/h_1 h = x/h$ . Since for Keynes the equal (or unequal) probabilities of hypotheses on evidence represent the equal (or unequal) degrees of confirmation or evidential support for those hypotheses, we may express his definition of irrelevance (and relevance) more exactly using Carnap's symbolism and notation for confirmation: evidence  $e_1$  is irrelevant to a hypothesis  $h$  on  $e$  if and only if  $c(h, e.e_1) = c(h, e)$ .

Keynes proceeds to use these concepts to elucidate the principle of indifference and it is worth our while to consider in full the passage in which he does this:

This distinction enables us to formulate the Principle of Indifference at any rate more precisely. There must be no relevant evidence relating to one alternative, unless there is corresponding evidence relating to the other; our relevant evidence, that is to say, must be symmetrical with regard to the alternatives, and must be applicable to each in the same manner. This is the rule at which the Principle of Indifference somewhat obscurely aims.

We must first determine what parts of our evidence are relevant on the whole by a series of judgements of relevance, not easily reduced to rule, of the type described above. If this relevant evidence is of the same form for both alternatives, then the Principle authorises a judgement of indifference [i.e. a judgement of equal probability]

The interesting feature about this line of argument is that Keynes, very much in contrast to Carnap and Carnap's interpretation of him, does not use the principle of indifference to define a measure function and corresponding confirmation function which would determine the degree of probability of some hypothesis  $x$  on evidence  $h$ ; rather Keynes requires the existence of judgments, similar to those determined by a confirmation function as a pre-condition for the employment of the principle of indifference. That is to say, Keynes' 'judgments of relevance' and judgments that the evidence is 'symmetrical with regard to the alternatives' and 'applicable to each in the same manner' which are the pre-requisites for employing the principle of indifference are all judgments as to the degree and means by which evidence supports, or confirms, the hypotheses in question. As such these are judgments that in a rigorous quantitative system such as Carnap's would be determined by the confirmation function. Whereas Carnap determined a measure function and confirmation function by using the principle of indifference, Keynes requires for use of the principle a series of judgments equivalent to those determined by a confirmation function, except that the judgments of relevance, irrelevance, symmetry and applicability that Keynes requires are, primarily, comparative rather than

numerical judgments of confirmation. (For example a judgment that  $h_1$  is relevant to  $x$  on  $h$  is a judgment that  $h h_1$  supports  $x$  to a greater or lesser, but not equal, degree as  $h$  itself supports  $x$ ).

How then should we characterize the logical status of the principle of indifference in Keynes account if it is not used to determine judgments of confirmation or evidential support, but rather presupposes them? Keynes in summarising his views on the principle provides us with the following answer:

In the first place we have stated the Principle of Indifference in a more accurate form by displaying its necessary dependency upon judgments of relevance and so bringing out the hidden element of direct judgment or intuition, which it was always involved. It has been shown that the Principle lays down a rule by which direct judgments of relevance and irrelevance can lead to judgments of preference and indifference. [judgments of unequal and equal probability]

What Keynes has accurately seen here is that the principle simply makes the truth conditions for judgments of equal and unequal probability to be judgments of relevance and irrelevance of evidence on the hypotheses. Like Keynes' less exactly defined but similarly important judgments of applicability and symmetry of evidence, these judgments (as we just noted) clearly concern the greater, lesser or equal degree to which various pieces of evidence support the hypotheses we are concerned with. Accordingly, Keynes's account of the principle turns out to be remarkably similar to that offered earlier, namely that the principle does nothing more than fix comparative judgments of evidential support as the truth conditions for judgments of probability.



Moreover, we find in Keynes exactly the same tacit assumption noted before, viz. that the evidence in relation to which we are to apply the principle is the total amount available. This much, I think, is clear from his insistence that in applying the principle there should be no relevant evidence relating to one alternative unless it is balanced by relevant evidence for the other alternatives. Only if the total available relevant evidence were considered could we determine if this were so - if we considered only part of the relevant evidence we could not know that some other part might be favourable to one alternative with no similar evidence in favour of the other alternatives. Thus in so far as Keynes recognized in his analysis of the principle that the truth conditions for the judgments of equal and unequal probability determined by it were comparative judgments of evidential support, where the evidence employed for the comparisons is the total relevant evidence available, his analysis wholly supports the conclusion reached before - that the principle of indifference simply fixes identity conditions for judgments of probability in terms of the comparative concept of evidential support, given that the evidence we are to use in comparing degrees of evidential support for our hypotheses is always to be the total relevant evidence available.

With this conclusion in mind we are now in a position, I think, to understand the long standing appeal of the principle and the great historical importance attached to it. This is most easily done, if we begin by considering the principle in so far as it specifies identity conditions for judgments of probability in terms of comparative judgments of evidential support. The additional fact that it is the total available evidence we use in making these comparisons is best left as a separate matter for later discussion.

Let us now recall that the discovery and assertion of the identity conditions for a concept is far from being an idle enterprise; following Frege it has been widely appreciated that statements of identity conditions for a concept fix the sense of that concept in that they express the conditions in virtue of which we repeatedly apply that concept. Thus any assertion that fixes identity conditions for a concept will serve as at least the rudiments of a semantic definition of that concept. In so far, then, as the principle of indifference specifies the identity conditions for judgments of probability between hypotheses in terms of comparative judgments of evidential support for those hypotheses, it serves as a rudimentary form of the definition of probability as the comparative concept of evidential support. Considering the widespread appeal of the definition of probability as a relation of evidential support and the well-known preference of traditional writers for comparative probability judgments, it is hardly surprising that an even rudimentary and inexplicit form of the definition of probability as the comparative concept of evidential support would play a central, even the central, role in the history of probability. The degree of thorough going conviction with which the principle was once widely advocated is now seen to stem not from its perhaps dubious claims for assigning 'a priori' probabilities but rather from its status as a tacit form of the highly plausible definition of probability as the comparative concept of evidential support.

Of course, it should be stressed that no proponent of the principle explicitly put it forward with the aim of defining probability as the comparative concept of evidential support. Indeed, in the classical theory of probability the principle was only used as a hand-maiden to the allegedly correct definition of probability as the ratio of favourable cases to possible cases, where the cases were equiprobable (or 'equipossible' as was sometimes said in a rather unsuccessful attempt to mask an obvious circularity). But as we have seen specifically with the passage from Jeffreys' Theory of Probability, much of the principle's charm has stemmed from at least an appearance of analyticity and this indicates that, at the very least, the principle should have been thought of as a direct consequence of some definition of probability. Having seen that the principle fixes identity conditions for probability judgments in terms of comparative judgments or, as we might now say, comparative relations of evidential support, we need only recall that statements that fix identity conditions for a concept determine the sense of that concept in order to see the principle as a form of the definition of probability as the comparative concept of evidential support.

To put this same point another way, my claim is that the principle merely fixes identity conditions for probability judgments in terms of comparative relations of evidential support and, whether or not it is recognized by anyone who does so, fixing identity conditions for a concept amounts to fixing a particular sense for that concept. My claim here about the principle is, as I remarked

before, analagous to Mach's claim about Newton's Third Law - Newton did not regard the Third Law as a definition of inertial mass, rather he took that concept to be clear enough and thought of the Third Law as a purely empirical statement. What Mach realised was that the Third Law, among other things, provided an operational criterion of when objects were equal or unequal in inertial mass and so acted implicitly as a definition of inertial mass. In just the same sense, I wish to claim that by fixing identity conditions for judgments of probability the principle constitutes implicitly and in an admittedly rudimentary way a definition of probability.

Further confirmation of this interpretation of the principle as a rudimentary form of definition can be found if we turn to some of Wittgenstein's remarks on meaning. For him the conditions actually used for repeated application of a concept are a far better indication of the meaning attached to that concept than are any explicit statements put forward as a definition - particularly if such statements are put forward by philosophers. There can, I think, be little doubt that, historically, writers who advocated the principle of indifference were not in the least lax in actually using (to the point of abusing) the conditions for equal probability expressed in the principle. Thus if they did not explicitly state that the principle of indifference was to be taken as a definition of probability, their extensive use of comparative relations of evidential support as criteria of identity for probability statements showed that they did in fact take probability to be defined as the comparative concept of evidential support.



Now, this interpretation of the principle of indifference can easily be objected to on the grounds that, as it stands, it provides no account of the primary (and perhaps exclusive) use to which the principle was historically put - the determination of unique quantitative probability values. Interpreting the principle as a version of the definition of probability as the comparative concept of evidential support seems, at first glance, to be at variance with this practice in two important respects: first of all, a salient feature of the definition of probability as the comparative concept of evidential support is that probability assignments to hypotheses will usually take the form of non-quantitative comparative judgments as to the greater, lesser or equal degree of probability between various hypotheses. Secondly, the other salient feature of the definition of probability as the comparative concept of evidential support is that numerous different orderings of a given set of hypotheses in respect of comparative degrees of probability will be possible, for these orderings will be relative to the body of evidence chosen, e.g. the probability of  $h_1$  and  $h_2$  relative to  $e$  might be the same, but relative to different evidence  $e'$ ,  $h_1$  might be more probable than  $h_2$ . But this variability of the comparative probability of hypotheses, which features prominently in the definition of probability as the comparative concept of evidential support, seems a far cry from the actual practice of proponents of the principle who, apparently, used it with the hope of arriving at unique (and indeed numerical) probability values.

The reply to this two-fold objection to my interpretation of the principle is, I think, highly instructive and in the end substantiates this

interpretation. First of all attention to the circumstances in which the principle of indifference was used to arrive at numerical results readily convinces us that its proponents did have a comparative conception of probability in mind. Almost invariably the principle was applied to sets of mutually exclusive and exhaustive alternatives, e.g. 'heads and tails', 'red and non-red' etc. In such cases (and only such cases) can comparative judgments of probability be turned into quantitative judgments. By the fact that the  $n$  alternatives are exhaustive we deduce that the sum of their separate probabilities is 1. If we then are able to make the comparative judgment that each of the  $n$  exclusive alternatives are equally likely, we can directly conclude that the probability of each is  $\frac{1}{n}$ . Thus the application of the principle in arriving at numerical results depends only on comparative judgments of probability, applied to members of an exclusive and exhaustive set of alternatives.

There now remains the apparent difficulty that in defining probability as the comparative concept of evidential support the principle of indifference would lead to different relational probabilities rather than the unique probabilities traditionally arrived at. Now it is widely appreciated that the definition of probability as a relation of evidential support, whether we use a quantitative or comparative concept, will yield unique results only if it is supplemented by a methodological criterion for selecting a unique body of evidence. And as we have already seen at length elsewhere in this

thesis the most common and natural proposal for selecting a preferred body of evidence on the basis of which we are to make unique assessments of probability is that we are to use the total available relevant evidence. In so far as the total available evidence is the body of evidence that is used in applying the definition of probability as the comparative concept of evidential support to given situations, this definition will yield unique orderings in respect of comparative degrees of probability. It then follows (by the above reasoning) that this definition will yield unique numerical probabilities in those situations in which our comparative judgment is that each of a set of mutually exclusive and exhaustive alternatives is equally likely in relation to the total available evidence.

Now, and this is the crucial point, the absence of 'reason for preference' on which alone the principle can determine equal probabilities consists in the total available evidence not supporting any alternative more than another. Thus, in point of fact, the principle of indifference is correctly applied to determine equal probabilities only on the basis of the total available evidence, for we can only legitimately judge that there is no reason for preference among competing alternatives when we have judged that the total available evidence does not support one more than the other. As we have just now seen, when the definition of probability as the comparative concept of evidential support is applied on the basis of the total available evidence it will yield unique and sometimes numerical probabilities. Thus the fact discerned earlier - but put off for full discussion until now - that the evidence which gives 'reason for

preference' when we apply the principle is the total available evidence insures that the principle will yield unique and sometimes numerical probabilities. The principle embodies the definition of probability as the comparative concept of evidential support in that it fixes identity conditions for the concept of probability in terms of comparative relations of evidential support; however that the principle is only correctly applied on the basis of the total available evidence insures that the relational definition of probability incorporated in it will have the unique application to actual situations that can lead to unique numerical probabilities.

In this analysis I have quite deliberately separated out in the principle components which, of course, are intertwined in its actual expression - the components being the principle's specification of identity conditions for judgments of probability in terms of the comparative concept of evidential support and the additional fact that the evidence used when the principle is correctly applied is the total amount available. Only by separating out these components can we, on my view, come to a proper understanding of the principle and its role in the history of probability theory. The principle, as I have claimed from the outset, is largely definitional in character for it specifies identity conditions for the concept of probability; unfortunately, the definitional aspect of the principle has long been overlooked for the definition of probability as the comparative concept of evidential support it offers is in a rudimentary form that, moreover, seems at variance with the unique and numerical probabilities so often arrived at by the principle. However, once we see the principle as a version of this definition and further realize that the principle is correctly applied only on the basis of the total available evidence, we see



that the principle can yield unique orderings of comparative degrees of probability. In the cases, so favoured by advocates of the principle, in which a number of mutually exclusive and exhaustive alternatives are found equally likely, we are led to our unique numerical results.

Admittedly this analysis hinges on my claim that the principle is only correctly applied on the basis of the total available evidence. In point of fact there does not seem to be any uniform interpretation among philosophers about the evidence on which the principle of indifference is correctly applied. Some critics of the principle, such as Kneale, appear to believe that its proponents intended it to assign equal probabilities on the basis of bodies of evidence corresponding to particular people's often deficient knowledge. Even writers who treat the principle more sympathetically sometimes interpret it as correctly applied on the basis of that knowledge of a particular person at a particular time. Mackie, who believes that the principle is instrumental in arriving at what he calls judgments of simple probability, says of the principle that 'its function is merely to indicate what is reasonable to believe in the light of whatever information we now have.'<sup>15</sup> This is also the way in which the principle is characteristically interpreted when referred to in decision theory.

However, another recent writer, Blackburn, analyzes the principle in such a way that it is only correctly applied on the basis of the total evidence one could reasonably be expected to gather. In criticizing Kneale's characterization of how the principle was understood by classical writers,

Blackburn asserts that the proper formulation of the principle as originally intended is that

- (4) Nothing that I know, nor rationally believe, nor should have discovered, favours A rather than B and nothing that I know, nor rationally believe, nor should have discovered, favours B rather than A.

entails

- (2) It is unreasonable to have other than the same degree of confidence in A and B.<sup>16</sup>

Blackburn, although he does not mention it, appears to have some weight of authority on his side. Bernouilli, in the same section of Ars Conjectandi in which he first enunciated the principle of indifference, states what Keynes calls his second maxim, that in determining probabilities:

Non sufficit expendere unum alterumve argumentum, sed conquirenda sunt omnia, quae cognitionem nostram venire possunt, atque ullo modo ad probationem rei facere videntur<sup>17</sup> [It is not sufficient to weigh just one piece of evidence but all of the evidence is to be sought out which can enter into our knowledge and seems in anyway to relate to the demonstration of the matter]

Thus it seems clear, particularly from the occurrence of the verb 'possunt', that Bernouilli would have regarded the principle of indifference as correctly applied only on the basis of the total available evidence, with available evidence understood as that which can be known in a situation, rather than that which just happens to be known.

It is exactly because of the divergent opinions of the evidence on which the principle is correctly applied that I have been content throughout this chapter so far to leave vague just what body of evidence I meant when I spoke of the total available evidence. As we saw in Chapters IV and V, and again in this chapter, available evidence can, be, and has been, understood in a number of different senses ranging from an objective sense to a subjective one in which the total evidence available to a person at a time is just what he knows at that time. By leaving it vague in which of these senses 'available evidence' was to be understood, it was possible to give an analysis of the principle that remains valid whether we think the principle is correctly applied on the basis of the kind of evidence envisioned by Blackburn and Bernoulli (the evidence which can or should be known) or, instead, on the basis of the kind of evidence envisioned by Kneale and Mackie, (the evidence which is actually known).

I think it must be clear how the analysis I have given of the principle works if it is thought that the principle is correctly applied on the basis of the kind of evidence envisioned by Blackburn and Bernoulli. I have claimed that the principle embodies the definition of probability as the comparative concept of evidential support and yet leads to unique numerical probabilities because it is only properly applied on the basis of a unique body of evidence, the total available at a given time. Since two perfectly natural (but strong) senses of 'available evidence' which so fix a unique totality at a given time are that which can be known (as in Bernoulli) or that which should be known

(as in Blackburn), their interpretations of the principle agree perfectly with mine and so accord with my analysis.

What happens to this analysis if we accept Kneale's and Mackie's claim that proponents of the principle intend it to be applied by people on the basis of their perhaps deficient knowledge? If, as they claim, the principle was used, and was intended to be used, to determine probability values on the basis of the limited knowledge of particular people at particular times, it is still the case that the principle would have been used, and, clearly, must have been intended to be used, on the basis of the total knowledge known to those people. To parallel what was said previously, if a person asserted that relative to what he knew there was reason to prefer  $q_1$  to, say,  $q_2$ , he would obviously be misleading us if by this he meant some small part of his knowledge favoured  $q_1$  while his knowledge in its entirety in fact favoured  $q_2$ . Reason for preference, even when relative to a particular person's knowledge, depends not on that knowledge taken partially but rather on the totality of that knowledge considered on balance.

If the principle is understood as used in relation to a particular person's limited knowledge, the fact that the principle would only be properly employed on the basis of his total knowledge suffices to sustain the analysis offered here. This analysis purported to show how the principle, by embodying the definition of probability as the comparative relation of evidential support,



could be used to arrive at unique numerical probabilities because it was correctly applied to a given situation only on the basis of one body of evidence. Now of course the total evidence known to any person at a given time is just as much a single unique body of evidence as the total 'available' evidence, in some strong sense of that word. For different people, of course, the total evidence known at a given time will be different but not so for one and the same person. Thus if a person always applied the definition of probability as the comparative concept of evidential support expressed in the principle on the basis of his current knowledge, he would always be led to what for him would be unique orderings for hypotheses in respect of comparative degrees of probability. When he found a set of mutually exclusive and exhaustive hypotheses which were equally well supported by the totality of his current knowledge, the definition of probability expressed in the principle would allow him to assign unique and equal numerical probabilities to each hypothesis. To put the point generally, the total but perhaps limited knowledge of a particular person at a particular time, just as much as the total evidence then available (in some strong sense of that word), constitutes a unique body of evidence; accordingly the relational definition I have claimed to be embodied in the principle of indifference will yield results which are unique for anyone who applies it on the basis of his total knowledge.

Thus, the logical analysis I have given of the principle as embodying a relational definition of probability and yet leading to unique numerical probabilities because it is in fact applied on the basis of unique bodies of evidence,

is essentially the same whether we regard the principle as properly applied in relation to the total evidence 'available' in a strong sense of this word or the total evidence known to a particular person at a particular time. Since the latter evidence can also be called the total 'available' evidence - in the weak subjective sense of that word we found Carnap using - it has previously been convenient to express my analysis of the principle in terms of a vaguely understood notion of 'available evidence', capable of different interpretation including this weak one. Indeed since my analysis of it is essentially the same no matter which of these interpretations is given to the principle, it is equally acceptable and more convenient to use even rougher terms such as 'the extant evidence' or 'our evidence' etc. to indicate the body of evidence - whether we think of it as the total 'available' in a strong sense of that word or in the weak subjective sense - in relation to which the principle is applied to arrive at unique and sometimes numerical probabilities. Accordingly until the end of the next chapter when finer distinctions become necessary I will adopt such usage as convenient.

Having now argued for the definitional character of the principle - and shown no matter what we take to be the evidence on which it is correctly applied the definition expressed in it can lead to 'unique' numerical probabilities - I should like now to conclude this chapter by indicating how this interpretation of the principle accords with the definition of probability explicitly recommended by the principle's first and most ardent advocates, the classical theorists.

The definition of probability explicitly offered by them was that of the ratio of favourable cases to possible cases, provided each case was equiprobable or equipossible. The principle of indifference was thought by them simply to provide a criterion for determining equiprobable or equipossible cases. Now construing the principle itself as a definition of probability yields the apparent anomaly that the classical writers had in mind two distinct definitions of probability.

But this anomaly is only apparent for the two different definitions here are in no way incompatible - indeed the two definitions answer roughly to what we now distinguish as an uninterpreted axiomatic definition of the concept of probability and a semantic definition for the concept. If we understand 'equipossible cases' as a primitive undefined term, the classical definition of probability as the ratio of favourable to equipossible cases constitutes an uninterpreted axiomatic definition of probability; the principle of indifference, which even in classical writings was understood as fixing a sense for the concept of equipossible cases, is then to be regarded as the semantic rule which gives an interpretation to the basic undefined term of the axiom system.

Carnap, despite his criticisms of the principle and, in my opinion, misappropriation of it to provide a measure function, admirably sets the stage for this characterization of the relationship between the principle of indifference and the definition of probability as the ratio of favourable to equipossible cases:

At the first look, the classical theory seems to be much stronger than the modern axiom systems; in other words stronger than a mere theory of regular c-function [the class of functions for which the axioms of probability hold]. And we find indeed many stronger theorems stated by classical authors. However, a closer examination shows that proofs for these stronger theorems make use, explicitly or implicitly, of the principle of indifference. The classical theory claims to give a definition for probability, based on the concept of equipossible cases. The only rule given for the application of the latter concept is the principle of indifference; since we know today that this principle leads to a contradiction, there is in fact no definition for the concept of equipossibility. In order to base the classical theory on a consistent foundation, we may proceed in the following way. We regard it, not as an interpreted theory as it was intended, but as an uninterpreted axiom system with 'equipossible cases' as undefined primitive term without interpretation. Then we take the classical definition of 'probability' based 'equipossible'. Thus this definition is here an uninterpreted axiomatic definition.<sup>18</sup>

Carnap's remarks here seem all the more compelling when it is realized that the classical definition of probability as the ratio of favourable to equipossible cases is distinctly similar to the modern axiomatic definition of probability as the measure of an attribute space (favourable cases) in a sample space (possible cases). Moreover one can hardly deny that the classical authors as mathematicians pursued with great insight the consequences of an axiomatic definition of probability.

Now, of course, any axiomatic definition of probability requires supplementation by a semantic definition, which today is given explicitly by rules of interpretation for the undefined terms of the axiom system. We can hardly expect the classical writers to have given a semantic



definition of probability by explicitly formulating rules of interpretation in the modern style for all the terms of their axiom system. However as 'equipossible cases' is the fundamental term of the axiom system of the classical writers, it might well be that the semantic definition classical writers had in mind became expressed in the rule of interpretation for this fundamental term, the principle of indifference. And indeed this is just what we have found already in this chapter - the entire discussion in this chapter so far has been to the effect that the principle of indifference in fixing identity conditions for judgments of probability in terms of comparative relations of evidential support actually embodies (in a rudimentary way) the highly plausible semantic definition of probability as the comparative concept of evidential support. Thus my claim that the principle of indifference is, essentially, a semantic definition of the concept of probability in no way conflicts with the classical theorists definition of probability as the ratio of favourable to equipossible cases. Rather once we see that this definition without a rule of interpretation for the fundamental term 'equipossible cases' is an uninterpreted axiomatic definition of probability, there is every reason to expect that a rule which provided an interpretation for it would constitute, in essence, a semantic definition of the concept of probability. The rule which classical theorists proposed for interpreting the term 'equipossible cases' was of course the principle of indifference and so our analysis of it as offering a semantics for the concept of probability is wholly in accord with the definition they explicitly offered in terms of the ratio of favourable to equipossible cases.

Moreover, in this light the concept of probability the classical theorists articulated does not appear to be the relic of a bygone era that so many now regard it as; rather the concept of probability they articulated is close approximation to a highly plausible and thoroughly modern one. Their theory presented an astonishingly fruitful axiomatic definition of probability in terms of the ratio of favourable to equipossible cases and moreover, with the principle of indifference, they supplied in rudimentary form the highly plausible semantic definition of probability as the comparative relation of evidential support to supplement their axiomatic definition.

Although, needless to say, I hope that this conclusion already seems correct to the reader, it is of course now necessary to consider the many criticisms that have been levelled against the principle. Only if these can be explained - and to some extent explained away - by the interpretation I have given here of the classical theory and the principle of indifference can it be accepted that the concept of probability they articulated is a close approximation to a plausible modern concept. In the next chapter then I will turn to the myriad objections that have been raised to the principle of indifference and will attempt to show how these objections can be understood as natural, but not insoluble, objections both to the semantic definition I have taken the principle to have offered and to the rudimentary way in which it does offer it.

Footnotes for Chapter VI.

1. Carnap, LFP, Section 12B.
2. Jeffreys, Harold, Theory of Probability, 3'd ed., Oxford 1961, pp. 33-34 and p.19.
3. Keynes, John Maynard, op. cit., p.51.
4. Von Mises, Richard, op. cit., p.75.
5. Keynes op. cit., p.85.
6. Kneale, W., op. cit., p.147.
7. Ibid., pp. 172-173.
8. Keynes, J. M., op. cit., Chap. IV. Unless otherwise stated all quotations from Keynes in this chapter come from Chapter IV of the Treatise.
9. Ibid., p.111.
10. Carnap, Rudolf, The Continuum of Inductive Methods, Chicago 1952, p. 39.
11. Ibid., p.40.
12. Ibid., p.38 and LFP p.565.
13. Carnap, R., LF, p.565.
14. Ibid.,
15. Mackie, J. L., op. cit., p.201.
16. Blackburn, S. W., op. cit., pp. 119-120.
17. Bernouilli, J., Ars Conjectandi, Basil 1713, p.215, cited Keynes, J. M. op. cit. p.313.
18. Carnap, R., LFP, p.343.

## CHAPTER VII

## The Principle and Its Critics

Objections to the principle of indifference are so numerous that any attempt to deal with all of them will, perforce, be rather lengthy. As I hope to analyze at least the most important objections to the principle from the perspective offered in the last chapter, I will later subdivide this chapter into several sections. But let me begin what will be a lengthy - and concluding - chapter of this thesis with consideration of a topic already discussed at some length: Carnap's claim to have settled much of the controversy over the principle by demonstrating that, when formulated to avoid any contradiction, it leads to two different measure and confirmations functions of which one pair is clearly preferable.

My primary objection to his analysis is just his contention that the principle leads to the  $m^+$  and  $m^*$  measure functions and thus to the  $c^+$  and  $c^*$  confirmation functions. As we saw in our discussion of Keynes's treatment of the principle, judgments about comparative degrees of evidential support (i.e. judgments of relevance and irrelevance) are required before the principle can be applied. The principle, as Keynes realized, only equates equal probability with equal evidential support and so can only be applied to yield probability values when we have already made comparative judgments about degrees of evidential



support. Now Carnap's confirmation functions lay down in quite rigorous fashion rules for correctly judging quantitative degrees of evidential support. Thus the principle could be used after such quantitative judgments were made, for it is easy enough to deduce comparative judgments from quantitative ones. But this is a very different matter from using the principle to derive a quantitative confirmation function, for rather than leading immediately to quantitative judgments of evidential support the principle requires a series of comparative judgments of evidential support before it can even be applied. Hence Carnap's use of the principle of indifference to arrive directly at the  $c^+$  and  $c^*$  functions for the atomic sentences of a language has the cart very much before the horse - we require rules for determining comparative degrees of confirmation for the atomic sentences of the language before we can apply the principle at all.

More particularly, as indicated before, I find Carnap's derivation of the  $m^*$  and  $c^*$  from the principle of indifference spurious, because, despite the apparent use of the principle as applied to structure descriptions rather than state descriptions, it would require a secondary application of the principle to each state description falling under a particular structure description to lead to the  $m^*$  function. Thus it is particularly clear that the  $m^*$  and  $c^*$  functions are not arrived at by a legitimate use of the principle as originally intended but only by an apparent use that coincides with Carnap's desire to find a plausible confirmation function.

Moreover, on serious reflection, Carnap's ascription of the  $m^+$  and  $c^+$  functions to Keynes and other previous proponents of the principle is

not very plausible. As Carnap himself remarks these functions are 'tantamount to the principle never to let our past influence our expectations of the future.' Such a principle can hardly be seriously attributed to the author of The Economic Consequences of Peace. More generally we must remember that most advocates of the principle were at pains to emphasize that judgments of equal probability between hypotheses about the future were only justified if experience gave no reason to choose between them. Given that 'experience' can only have taken place in the past, such a reservation concerning the conditions for applying the principle clearly indicates that, contrary to Carnap's assertion, the proponents of the principle did believe that the past should influence our expectations for the future. Thus the ascription of the  $m^+$  and  $c^+$  functions to various proponents of the principle seems seriously at variance with their own application of it.\* As with the  $m^*$  and  $c^*$  functions, the  $m^+$  and  $c^+$  functions are not derived from the principle of indifference's traditional employment but rather are derived only by an apparent application of it in Carnap's own system.

Despite what I think is his misinterpretation of the principle as traditionally advocated, Carnap is, on the whole, obviously sympathetic to it. Not so with others who have discussed it - as we saw at the start of the previous chapter - and we should now consider these negative responses to it. To do this properly we must first elaborate in some detail on the exact nature of the

---

\*In fairness to Carnap, we should note that the ascription of the  $m^+$  and  $c^+$  functions to Wittgenstein, as opposed to Keynes and others, is not implausible. Wittgenstein in the Tractatus apparently does advocate giving equal measures to all elementary propositions and this is usually interpreted along the lines Carnap advocates.

evidence which offers no reason for preference among the alternatives which the principle, then, treats as equally probable. It will be my contention that much of the criticism that the principle has received depends on quite natural, but, in many cases, not insurmountable, misgivings concerning the nature of this evidence.

Now, the principle of indifference states that we are to hold hypotheses equally likely if there is no reason to choose between them. Further we saw this meant that the extant evidence gave no reason to choose between them, for reasons for preference were always derived from evidence. But, and I think this accounts for a great deal of the controversy surrounding the principle, there is an ambiguity as to how the extant evidence might give no reason to choose between alternative hypotheses. Since it is only evidence which can provide reason for hypotheses, one possible circumstance in which the extant evidence would give no reason to choose between our hypotheses is when that evidence is distinctly relevant to the matter in hand but, on balance, is equally favorable to each hypothesis. The clearest examples of such relevant evidence that is equally balanced between various alternatives are those in which we have information that the alternatives have in the past occurred, or will in the future occur, in equal proportions. To put roughly what has been discussed elsewhere more rigorously, consider a coin which has been carefully examined and found to be exactly symmetrical in its weight distribution. Furthermore, extensive tests on it and similarly weighted coins have revealed a percentage of roughly 50% heads in tosses by the kind of mechanism we are employing; we may also imagine we have extensive knowledge of the workings of our tossing mechanism. Here we very obviously have evidence relevant to the



competing hypotheses that on a given toss the coin will land heads or tails - indeed no information could be more relevant - but it should be clear that our evidence here is plausibly thought of as equally balanced between the two hypotheses supporting the hypothesis of heads no more or less than it supports the hypothesis of tails.

In terms of the analysis given in Chapter V of the relationship between frequency and evidential support we may, to put it generally, say that information concerning the occurrence of properties  $\phi$  and  $\psi$  in a particular trial constitutes the evidence pertaining to that trial, while our best estimates of the limit of relative frequency of different alternatives  $a, b, c$ , etc. in trials in which  $\phi$  and  $\psi$  occur determines the degree of support we regard such evidence as giving to hypotheses about the occurrence of  $a, b, c$  etc. in a trial. Thus statistical data to the effect that a series of trials held under conditions  $\phi$  and  $\psi$  result in equal proportions for the alternatives  $a, b, c \dots$  will allow us to believe with equal confidence in the hypothesis of alternative  $a$  as the outcome in a particular trial held under conditions  $\phi$  and  $\psi$ , as in the hypothesis of  $b$  as the outcome, and so on for  $c, d$ , etc.

But in contrast to such circumstances in which the extant evidence gives no reason for preference among alternatives because it is relevant and provides an equal measure of support for each, we have a more problematic circumstance in which our evidence may be said to give no reason for preference among alternatives. This more problematic circumstance occurs when the extant evidence is wholly irrelevant to the question at hand and, as such,



tells us nothing about any of our alternative hypotheses. This total absence of knowledge', so often referred to in the principle of indifference's history, is also obviously one in which our evidence gives no reason for preference among our alternatives, for such reason could only be provided by a body of evidence relevant to the matter which supported one alternative more than another.

Now before I go on to examine the consequence of the ambiguity that evidence may give no reason for preference either by being relevant and equally balanced or by being the null relevant evidence (as I will now call the latter circumstance), we should briefly look at least one passage from the literature surrounding the principle to convince ourselves that this ambiguity does exist, for, to my knowledge, it has never been explicitly articulated before.

In point of fact a crucial passage from Keynes, already cited twice, illustrates well this ambiguity:

There must be no relevant evidence relating to one alternative, unless there is corresponding evidence relating to the other... This is the rule that the Principle of Indifference obscurely aims at.

But quite clearly either the null relevant evidence or a body of evidence which is symmetrical, i.e. evenly balanced, in respect of the alternatives would fulfill this condition for the application of the principle.

Now admittedly not all proponents of the principle of indifference have employed it in a manner that is ambiguous between assigning equal probabilities on the basis of null relevant evidence or equally balanced relevant evidence - Carnap, for the purpose of constructing a priori measure functions, was exclusively concerned with the principle's application in relation to the null relevant evidence. However, this ambiguity does attach to the principle for the majority of writers on it, and has been dwelt on at length, because, taken in light of the definitional character of the principle brought out already, an appreciation of this ambiguity permits a clear understanding of much of the controversy surrounding the principle. With the principle construed as offering a definition or semantics for the concept of probability in terms of comparative relations of evidential support, the ambiguity outlined above can be seen to stem from fundamentally distinct features of the semantics offered by the principle.

To put this more clearly we should note that the comparative relations of evidential support with which the principle of indifference defines probability will hold between pairs consisting of evidence sentences coupled with hypotheses (which we also take to be sentences). For example, if we say that  $e$  gives an equal degree of evidential support to  $h$  as  $e'$  does to  $h'$  we have  $\langle\langle h, e \rangle, \langle h', e' \rangle\rangle$  as an element in the extension of the relation of equal evidential support. With this in mind we may characterise the two circumstances in which we have seen the principle to be applicable thus: in cases where our evidence is relevant to the question at hand, the identity

conditions for judgments of probability given by the principle are comparative relations of evidential support between ordered pairs of sentences  $\langle h, e \rangle$  and  $\langle h', e' \rangle$  where  $e$  is relevant to  $h$  and  $e'$  relevant to  $h'$ . When the evidence is the null relevant evidence, the identity conditions for judgments of probability given by the principle are comparative relations of evidential support between ordered pairs  $\langle h, e \rangle$  and  $\langle h', e' \rangle$  where  $e$  is irrelevant to  $h$  and  $e'$  irrelevant to  $h'$ . Since, as we have seen, the stipulation of identity conditions constitutes at least the rudiments of a definition or semantics for a concept, we may say that the ambiguity, as to the circumstances in which the principle of indifference applies, reflects the basic fact that the semantics offered by the principle includes comparative relations between ordered pairs of evidence and hypothesis where, in some cases, the evidence is relevant to the hypothesis in question, while in others it is not.

Cast in this form it is possible to understand, without necessarily accepting, the rationale that led eminent writers to argue that the principle could assign equal probability values in the total absence of relevant information. Consider any two pairs of sentences  $\langle h, e \rangle$  and  $\langle h', e' \rangle$  where  $e$  is irrelevant to  $h$  and  $e'$  irrelevant to  $h'$  ( $e$  and  $e'$  can be, but needn't be, identical). The character of the evidentiary relation (as we may call it) between the hypothesis  $h$  and its null relevant evidence  $e$  is obviously essentially similar to that of the relation holding between the hypothesis  $h'$  and its null relevant evidence  $e'$ . Now if we wish to construct a non-arbitrary semantics of probability on the basis of relations between evidence and hypothesis,

it is clearly desirable that all pairs of evidence and hypothesis bearing a similar relation to each other be treated in a similar fashion. This means that if we are to assign any probability value to these pairs of evidence and hypothesis, we must assign equal probabilities to each pair. Here the assignment of equal a priori probabilities springs from a desire to avoid arbitrariness and this accords with the views expressed by a number of writers in justifying equal a priori probabilities - for example Jeffreys remarks, already cited, on the need to assign equal a priori probabilities to avoid 'prejudice'.

To put the matter generally, we see that in giving the semantics for the concept of probability in terms of comparative relations of evidential support, we are naturally faced with a decision at some point as to whether we can compare, in respect of degree of evidential support, pairs of evidence and hypothesis whether the evidence is the null relevant evidence. Once decided the need for a non-arbitrary semantics requires that all such pairs be treated in the same way. Now, in fact, we here have two options: on the one hand we may wish to admit that such pairs can be compared in degree of evidential support and this decision then leads to the conclusion that all such pairs are equal in this respect. On the other hand, we may simply decide that such pairs are not comparable in degree of evidential support and, again, treating them all in a similar fashion, we would simply refuse to assign comparative degrees of probability to any two of such pairs. That is to say, in giving our semantics for the concept of probability in terms of comparative relations of evidential



support, we may simply decide to exclude from the semantics comparative relations between pairs of evidence and hypothesis where the evidence is wholly irrelevant. This, of course, does not stop us from retaining the other and, I think clearly more plausible part of the semantics offered by the principle of indifference, the part which defines probability in terms of comparative relations of evidential support between pairs of evidence and hypothesis where the evidence is relevant.

Now, of course, it might be argued that if we were to reject that part of the principle's semantics that includes and treats as equal each evidentiary relation between a hypothesis and its null relevant evidence, one has, so to speak, thrown the baby out with the bath water. That is, historically much interest in the principle has focused on its capacity to assign a priori probabilities and, of course, the discovery of a means to assign a priori probabilities is not unlike the discovery of the proverbial philosopher's stone. But to hold that the principle's main interest and appeal stems from its determination of a priori probability values is, I think, to get the matter the wrong way round. Once we see the principle as offering a semantics for the concept of probability in terms of comparative relations of evidential support and realize how plausible this is in those cases where the evidentiary relations to be compared are those between hypotheses and their relevant evidence, we can see the comparison in the cases where the evidence is irrelevant as a natural, but misguided, expansion of a plausible semantics to include less plausible cases.

Now as the principle of indifference is genuinely ambiguous between assignments of equal probability in relation to evenly balanced and relevant evidence and the null relevant evidence, the principle must be construed as offering semantics including both sorts of cases. But once this is appreciated we can see that the more plausible part of the semantics offered by the principle would attract us by having that appearance of analyticity found in plausible definitions, while the less plausible part would puzzle us and seem objectionable. This, I think, largely explains why the principle has appeared at times almost analytically valid and other times highly objectionable, for the objectionable features stem from just one part of a semantics which is, in other parts, highly plausible.

Indeed it is just because much of the controversy over the principle can, in my opinion, be traced back to this ambiguity that I have devoted so much attention to elucidating it in a chapter ostensibly concerned with the principle's critics. Once the nature and consequences of this ambiguity are appreciated it is possible, in large measure, to deal satisfactorily with most objections to the principle. In fact in what remains of this chapter I will try to show explicitly, first, that all but one of the most common objections to the principle are readily avoided if the principle's use is restricted to cases in which we have relevant evidence and, secondly, that there is every reason to believe that most classical theorists actually envisioned the principle to be applied in such cases where relevant evidence existed.

As to the first point, it is clear, for example, that restricting the principle's use to cases in which there exists relevant evidence would overcome the objection that, in violation of the maxim 'ex nihilo nihil', the principle produces knowledge out of nothing, for it is only in the total absence of relevant evidence that we know nothing concerning a particular set of hypotheses. Similarly such a move would forestall Von Mises' objection that if 'we know nothing about a thing, we can not say anything about its probability', for restricting the principle to circumstances in which the evidence was relevant to the matter at hand would insure that 'we knew something about a thing' before we assessed its probability.

In the same vein, it is interesting to note that if recast to assign equal probabilities solely on the basis of equally balanced relevant evidence, rather than on null relevant evidence as well, the principle will actually be in a form quite similar to an apparently fundamental modification of the principle urged by many of its critics. Kneale, as evidenced by the passage cited at the start of the previous chapter, claimed that a correct basis for probability judgments was not the 'absence of knowledge' but the 'knowledge of absence'. In a similar vein Reichenbach, while criticizing the principle, is prepared to accept a criterion of equiprobability if 'the equiprobability does not appear as following from the absence of reasons, but as a result of the existence of definite reasons'<sup>1</sup>. Mellor also attempts a reconstruction of the principle along these lines<sup>2</sup> and from Keynes's remarks on the history of the principle it would appear that Czuber attempted such a reformulation of the

principle in the 19th Century under the guise of the 'Principle of Compelling Reason'.<sup>3</sup>

Now the existence of relevant but evenly balanced evidence that I have indicated as the more satisfactory grounds for the principle of indifference's assignment of equal probabilities does constitute just such a definite, or compelling, reason for holding alternatives equally probable. We have seen how definite reason for assigning unequal probabilities could only be provided by relevant evidence that supported one hypothesis more than another and, so, by similar reasoning, it is clear that what is meant by definite reason for assigning equal probabilities would be the existence of relevant evidence that is equally balanced between hypotheses. Since just such evenly balanced relevant evidence can be used in connection with the principle we see that, restricted to such evidence, the principle is in a form not merely acceptable to, but actually advocated by, many of its critics.

In addition I think it should be pointed out in this context that in practice very little would be lost by abandoning the principle's use in determining a priori probabilities for, in reality, we rarely, if ever, are faced with hypotheses about which we have absolutely no knowledge at all. Realistically speaking, then, the determination of a priori probabilities is not a practice so necessary that we need severely lament its passing.



### The Contradictions

I have closed the above section by remarking that many objections to the principle can be avoided - at little cost - by abandoning its application in relation to the null relevant evidence but retaining it in cases where relevant evidence exists. Significantly enough, it can, in line with this position, be shown that the contradictions to which the principle is known to lead only occur when the principle is applied in relation to the null relevant evidence, rather than to a body of evenly balanced relevant evidence. Consider Keynes's example of the colour of a particular book. Keynes explicitly states that the example is one in which we have 'no evidence relevant to the colour' of the book in question, so it only remains to be seen that this contradiction could not have arisen if, instead, we had a body of evenly balanced relevant evidence. Now, the first step of the reasoning that leads to our contradiction is that having no reason to choose between 'This book is red' and its negation, we must assign each a probability of  $\frac{1}{2}$ . This first step could be arrived at by an application of the principle of indifference in relation to an evenly balanced body of relevant evidence - for example it might be known that red, being an eye-catching colour, was used very nearly half the time by publishers, with the remaining publications divided evenly between a number of common colours such as blue, black, green etc. If that were the total available evidence we would have no reason to prefer red to not red or vice versa and so, by the principle of indifference, each would be  $\frac{1}{2}$  probable. But now we can not take the next step in our reasoning to a contradiction - that we have no reason to prefer the hypothesis 'This book is black' to its negation - for we do indeed have a reason to prefer its negation as the predicate 'not black' encompasses the more frequently used colour red.

It should be clear that a similar line of argument will suffice to deal with many of the celebrated contradictions that the principle of indifference leads to: if we take a body of equally balanced relevant evidence as the basis for our judgment that there is no reason to choose between a set of mutually exclusive and exhaustive alternatives, that body of evidence will provide reason for preference among the members of those different sets of mutually exclusive alternatives that exhaust the same problem, whose use leads to contradictions. The alternatives under the second set will encompass varying numbers of alternatives from the first set and, by hypothesis, our extant evidence gives equal reason to expect only equal numbers of the alternatives from the first set.\*

---

\*I am indebted to Mr. Mackie for indicating to me in a public discussion of a version of my argument here a point that may require some elaboration. He made the suggestion that the assignment of equal probabilities on the basis of equally balanced relevant evidence, as I have described it, apparently does lead to absurdities, if not contradictions. To keep with the example of the colour of a book, consider someone who has observed only a limited number of books, say, 20. Half of these have been red, the other half not red, and, in fact, these other 10 have all been black. All the evidence he knows is equally balanced between the hypothesis of red versus not red and so he concludes, rather absurdly we should think, that red is as likely as not. Mr. Mackie added that I could not in all fairness rule out such absurdities by suggesting that the man had simply failed to uncover the extant relevant evidence that book publishers use other colours; if I could appeal to this fact in explaining away such absurdities, then (Mr. Mackie argued) one should have been allowed to appeal to such relevant evidence to avoid the contradiction of Keynes's initial example. The point, however, is that Keynes's initial example concerns the assignment of probabilities in the absence of all relevant evidence and so - if we are to remain within his terms of reference - we can not appeal to relevant evidence to avoid the contradiction. That such absurdities and contradictions can be avoided if we assign probabilities on the basis of relevant evidence is of course something I would admit and indeed this is at the heart of my analysis of the contradictions that principle leads to.

Once it is accepted that I, unlike Keynes in his initial example, can appeal to the existence of relevant evidence to remove the appearance of absurdity, the difficulty raised by Mr. Mackie is easily dealt with. Interestingly enough what Mr. Mackie points to is an absurdity not an outright contradiction, as in Keynes's example. I think it is clear that the absurdity of concluding after observation of 10 red and 10 black books that red is as likely as not is really the absurdity of reaching conclusions about probability after very limited observations. When I spoke of evidence equally balanced between different alternatives in virtue of the equal proportionality found in actual trials, I took it for granted that we had observed an extensive number of relevant trials. Unless we have done so, we can be led into absurdity - but not contradiction - by the method of weighing evidence I have relied on because this method is appropriate only when extensive observations have been made.

The contradictions in regard to geometrical probabilities can be dealt with in a similar, but slightly more complex, way. As an example let us consider Bertrand's paradox, the most famous of the difficulties that the principle gives rise to in the field of geometrical probabilities. In this paradox we are asked to determine the probability of a chord in a circle drawn at random exceeding the length of the side of an inscribed equilateral triangle. If we think of the chord as fixed by determining an angle from a tangent to the circle, the principle of indifference apparently yields the conclusion that the probability of the chord exceeding the length of the side of the triangle is  $\frac{1}{3}$ : the chord will be longer than the side when, and only when, the angle from the tangent is in the range  $60^\circ$ - $120^\circ$  and having no reason to prefer the angles of  $60^\circ$ - $120^\circ$  to  $0^\circ$ - $60^\circ$  or  $120^\circ$  to  $180^\circ$ , we conclude by the principle that each of these possibilities equally likely. Alternatively we may think of the chord as fixed by the location of its mid point. Viewed in this fashion the principle of indifference can yield the result that it is  $\frac{1}{2}$  probable that the chord will be longer than the side of the triangle: if, and only if, the mid-point of the chord is at a distance less than  $\frac{r}{2}$  from the center will the chord be longer and since there is no reason for the chord's mid-point to be less than  $\frac{r}{2}$  from the center, rather than more, each of these possibilities is equally likely. But, to give us a third application of the principle to this problem, it should be clear that if the chord is longer than the side of the triangle when and only when its mid-point is less than  $\frac{r}{2}$  from the center of the circle, a concentric circle of radius  $\frac{r}{2}$  encompasses all the area in which the mid-point of the chord can fall if it is to be longer than the side of the triangle. Since the area of this concentric circle is  $\frac{1}{4}$  that of the



original circle and since there is no reason to expect the mid-point to fall within any particular area, the principle of indifference here yields the result that the probability of the chord being longer than the side of the triangle is  $\frac{1}{4}$ .

The problem here, as Keynes acutely saw, is that the principle of indifference is applied to choices between three different infinite sets of possibilities where every possibility in each set gives a determinate result to the question at hand. That is, selection of an angle from the possibilities  $0^{\circ}$ - $180^{\circ}$  determines the relative lengths if we think of the chord as drawn at that angle to a tangent. But similarly, selection of a linear distance from the center for the mid-point of the chord determines the relative lengths. And finally selection of the mid-point's location within the area of the circle determines the relative lengths. Now the principle of indifference, in the absence of any relevant information, is said to accord each possibility within the three different sets an equal measure i.e. each angle, distance and area are to be treated equally, and this in turn leads to inconsistent results.

Accordingly, as Keynes recognized, it is a necessary condition for the resolution of Bertrand's paradox that one particular set of possibilities be decided on as the appropriate one. This, as Kneale's later and more detailed analysis of the difficulty suggests,<sup>4</sup> is easy enough if we know the specific means by which the chord is to be drawn. If we know that the location of its mid-point is to be determined by the fall of a rain drop, it seems most appropriate to determine the probability of the chord exceeding the side in length by reference to the area in which the mid-point lies and, if we take all areas as equally likely, our result is  $\frac{1}{4}$ . Alternatively, if we know that the chord is drawn along the line of a pointer



that is spun around an axis on the edge of the circle, it is most appropriate to determine the probability value by reference to the angle formed by the chord to a tangent, and, if we hold each angle equally likely, our result is  $\frac{1}{3}$ . Other mechanisms for drawing the chord would make the distance of the mid-point from the center the appropriate factor to be considered.

The important thing to note here is that the line of a solution to the paradox - the details of which can be found in Kneale's discussion cited above - requires us to have knowledge of the method by which the chord is actually drawn. But once we have such knowledge we have passed from our state of total ignorance and have begun to acquire evidence which is relevant to the problem at hand. So, as with the case of non-geometric probabilities, the difficulties that the principle give rise to can be resolved by an appeal to relevant evidence.

This perhaps becomes clearer if we consider that our evidence concerning the mechanism by which the chord is drawn is not, in itself, sufficient to determine the unique probability value we desire. For example, given that we know the chord to be drawn by spinning a pointer aligned with the edge of the circle, we only know that the probability of the chord exceeding the side in length depends on the probability of the pointer forming various angles with its axis. Only if we take all angles equally likely do we arrive at the result of  $\frac{1}{3}$ .

Here, of course, we might try appealing to the principle of indifference-knowing nothing about the nature of the spinning mechanism,

we might assume every angle equally likely, as there is no reason to prefer any particular angle. But this is nothing more or less than an application of the principle of indifference to a problem of geometrical probabilities in the absence of relevant evidence and, as Bertrand's paradox shows in a particularly elegant and appealing way, such applications lead to contradictory results. As Keynes puts it, the basic problem with applying the principle to geometrical probabilities is that

In general, if  $x$  and  $f(x)$  are both continuous variables, varying always in the same or in the opposite sense, and  $x$  must lie between  $a$  and  $b$ , then the probability that  $x$  lies between  $c$  and  $d$ , where  $a < c < d < b$ , seems to be  $\frac{d-c}{b-a}$  and the probability that  $f(x)$  lies

between  $f(c)$  and  $f(d)$  to be  $\frac{f(d) - f(c)}{f(b) - f(a)}$ . These expressions, which represent the probabilities of necessarily concordant conclusions, are not, as they ought to be, equal.<sup>5</sup>

So although we have evidence relevant to the selection of a chord i.e. knowledge that a spinning mechanism is used, and thus can avoid Bertrand's paradox, we will need further relevant evidence about the nature of the spinning mechanism to avoid arriving at contradictory conclusions about it through use of the principle of indifference.

In point of fact most serious evaluations of Bertrand's paradox have indeed been forced eventually to call on further relevant evidence to establish the equal probabilities required for a numerical answer to the original problem. Kneale, for example, maintains, by use of a celebrated argument of Poincaré,

that only a few further plausible assumptions, such as the relative size of the areas covered by each possibility, are required for an ultimate unique solution. Meller contends in reply that the assumptions are actually quite sweeping.<sup>6</sup> In any event, both writers agree that the paradox requires for its solution further empirical evidence which will be highly relevant to the question at hand. Moreover, in so far as we are able, on the basis of such relevant evidence, to reach the conclusion that a certain set of alternatives are equally likely, this evidence can be said to be equally balanced in favour of each alternative. As with non-geometric probabilities the essential grounds for our evidence to be equally balanced will be the occurrence of the alternatives in equal proportions in the past. If they or similar events did not occur in the past we would have no relevant evidence, and if they occurred in unequal proportions, the evidence would favour some alternatives over others. In fact, it seems fairly clear it is just such observed equal proportions on actual occasions involving geometrical probabilities e.g. spins of roulette wheels, that warrant the assignment of equal probabilities to each alternative.

Thus in considering geometrical probabilities we find that the contradictions to which the principle can lead only arise with its use in relation to the null relevant evidence and are naturally resolved by an appeal to some body of relevant evidence. Furthermore we see that in the actual probabilistic situations in which we might use the principle, the existing empirical data is relevant to the problem and can be shown to be equally balanced in the very important sense that it indicates the equal proportionality of the results obtained for each alternative. So in bringing out emphatically the difficulties involved in applying the principle

in relation to the null relevant evidence, as well as pointing out the efficacy of the principle in relation to relevant but evenly balanced evidence, a consideration of geometrical probabilities strongly supports the suggestion that the principle can be satisfactorily formulated by restricting it to cases in which the available evidence is relevant and evenly balanced.

Now, as already indicated, such a restriction on the principle's application amounts, on the interpretation offered here, to a decision to exclude comparative relations involving the null relevant evidence from the semantics given for the concept of probability. Curiously enough, the contradictions we have been discussing which recommend such an exclusion are themselves engendered by the need for a non-arbitrary semantics. Or so, at least, I have claimed, for it was argued above that the conclusion that any two pairs of hypotheses and their null relevant evidence must be judged equal in degree of evidential support (if they are to be compared at all) in order to avoid arbitrary treatment of essentially similar pairs. And it is, of course, just this conclusion as to the equal degree of evidential support of any two such pairs that leads to the contradictions when the equal degrees of evidential support are identified, by the principle of indifference, with equal degrees of probability. It is now possible to give a perspicuous account of the cause of these contradictions - to do so we should note that the contradictions in question are in the first instance contradictions of the axioms of probability. For example Keynes's 'impossible case of three exclusive alternatives all as likely as not', as he expresses the difficulty with the colour of a book, is impossible because it



violates the axioms, or syntax, of probability. By the special addition theorem we know that the probability of a disjunction of exclusive hypotheses is the sum of their individual probabilities and so, on one line of reasoning, the probability of the hypothesis that a given book is red or black or blue would be  $1\frac{1}{2}$ . This contradicts the axiom that all probabilities values are real numbers in the interval 0-1. Other contradictions violate this or other axioms, e.g. Bertrand's paradox violates the axiom, formulated for a two place probability function, that there is only one probability value for a term  $b$  in relation to a constant term  $a$  (taking  $a$  as the null relevant evidence and  $b$  as the hypothesis that the chord exceeds the side of the inscribed triangle in length).

Put generally then, the difficulty involved in the contradictions is this: If we desire a non-arbitrary semantics for the concept of probability in terms of comparative relations of evidential support and wish, moreover, to include in it comparative relations involving the null relevant evidence, we are required to stipulate that any two hypotheses in relation to their null relevant evidence are equal in probability, for each such pair of hypothesis and evidence is an instance of the same evidentiary relation. Unfortunately there exist hypotheses drawn from different sets of mutually exclusive and exhaustive hypotheses which can be compared each one by one and judged equally probable in relation to the null relevant evidence. But taken collectively the probability values so determined violate the axioms or syntax of probability.

If this characterization of the contradictions to which the principle can lead is correct, we might expect to find arbitrary any solution to them other

than abandoning all applications of the principle on the basis of the null relevant evidence, for it was just the desire for a non-arbitrary semantics including cases of null relevant evidence that led to the contradictions. This, in fact, is just what we do find with the various proposals that have been put forward to resolve the contradictions without wholly abandoning the principle's a priori application. Ayer<sup>7</sup>, citing an example of Watling<sup>8</sup>, puts just this point in general and compelling fashion. Here I shall concentrate on the particular features of the one line of solution to the contradictions we have already considered, viz., Carnap's restriction of the a priori applications of the principle to Q-divisions.

As we saw in the previous chapter, for a simple example of a language with one family of predicates ('W', 'R', 'B') and two individual constants ('a' and 'b'), Carnap's Q-division is the three predicates  $Q_1 = R$ ,  $Q_2 = W$  and  $Q_3 = B$ , for these form a division and involve explicitly all the primitive predicates of the language. Carnap held it to be a matter of controversy whether, in applying the principle to the Q-division, we held individual or statistical distributions equally probable, with sound inductive reasoning favouring the latter. Our concern for the moment, however, is the selection of the Q-division as the one to which the principle is to be applied and we shall assume for the sake of expository simplicity that once we have chosen a particular division, we will apply the principle to individual distributions of that division. As we saw, such an application of the principle to a Q-division leads to equal probabilities for each state description, for an individual distribution under the Q-division is a state description. But an alternative division was available in our example with the so called molecular predicates  $M_1 = W$  and  $M_2 = R \vee B$ , or more

simply  $W$  and  $\sim W$ . Using the principle to assign equal probabilities to individual distributions under this division led to quite different probability assignments than if we applied the principle to the  $Q$ -division, and, to avoid such a contradiction, we are required by Carnap to choose the more 'natural'  $Q$ -division. This, in particular, meant that we were to hold all state descriptions equally likely and thus we would, for example, deem the hypothesis ' $Wa$ ' less likely than ' $\sim Wa$ ' for the latter holds in 6 state descriptions (' $Ra. Wb$ ', ' $Ba. Bb$ ' . . . .), while the former only holds in three (' $Wa. Wb$ ', ' $Wa. Rb$ ', ' $Wa. Bb$ ').

But can the greater degree of probability assigned to ' $\sim Wa$ ' be justified in the absence of all relevant evidence? If, by the principle's dictum that we are to treat alternatives as equally likely if and only if there is no reason to choose between them, the unequal probability values assigned to ' $Wa$ ' and ' $\sim Wa$ ' would require a reason to be known in favour of ' $\sim Wa$ ', but, in the absence of any relevant evidence, there can be no such reason. Only relevant evidence can provide a reason for preference.

Now it has been claimed as early as Sigwart,<sup>9</sup> and more recently by Keynes and Blackburn<sup>10</sup>, that in such cases we do have knowledge that provides a reason for preference between, in our example, ' $Wa$ ' and ' $\sim Wa$ '. The knowledge is that already discussed, namely that ' $\sim Wa$ ' involves the assignment of two primitive predicates to  $a$ , and thus ' $\sim Wa$ ' holds in 6 individual distributions under the  $Q$ -division, whereas ' $Wa$ ' involves the assignment of one and thus only holds in three individual distributions under

the Q-division. In Carnap's terminology what is claimed here is that we always have relevant evidence which favours hypotheses covering more individual distributions under the Q-division (e.g. ' $\sim$ Wa') over those that cover less ('Wa'). That the existence of such relevant information could resolve some of the contradictions fully accords with the point made already, that the contradictions only arise in relation to the null relevant evidence. But if we claim that this knowledge does provide reason for preference we must abandon any pretence that our probability values are determined in the total absence of relevant evidence, for we now are assigning probability values on the basis of the evidence, deemed relevant, of the number of primitive predicates covered by a hypothesis - which of course is proportional to the number of individual distributions the hypothesis holds in under the Q division. Given that we deem as relevant evidence the number of primitive predicates involved in a hypothesis and thus as relevant the number of individual distributions under the Q-division, the assignment of unequal probabilities to 'Wa' and ' $\sim$ Wa' is not done in the total absence of relevant evidence, but rather is done on the basis of relevant evidence favourable to one hypothesis.

Put in this way I hope it begins to appear somewhat suspicious to claim that the number of primitive predicates, and hence individual distributions under the Q-division, is evidence relevant to the probability of hypotheses of our language. As far as I can see, in saying this we have not offered an argument of any kind, but have only stipulated that we intend to regard the number of individual distributions under the Q-division as relevant evidence. What



Independent rationale can be given for regarding this as relevant evidence? Can we not just as easily regard the number of individual distributions under the division formed by the predicates  $M_1$  and  $M_2$  as the relevant evidence which decides the matter. We may feel like saying that hypotheses covering more distributions under the Q-division can happen in more ways; after all ' $\sim Wa$ ' can occur as ' $Ra$ ' or ' $Ba$ '. But only by regarding the matter from, so to speak, the perspective of the Q-division can such hypotheses happen 'in more ways'; it is equally true that ' $\sim Wa$ ', like ' $Wa$ ', happens in only one way, namely by a not being white.

Again one might be tempted to reply here that hypotheses like ' $\sim Wa$ ' can really happen in more ways than ones like ' $Wa$ ', because there are more state descriptions in which ' $\sim Wa$ ' holds and the number of state descriptions in which a hypothesis holds is the important fact. But now we have come full circle: a state description simply is an individual distribution under the Q-division and so claiming that the number of state descriptions in which a hypothesis holds is the relevant evidence, we simply have re-stated our previous claim that the number of individual distributions under the Q-division is the relevant evidence. And again I ask why is this evidence, as opposed to evidence of the number of individual distributions under some other division, the relevant evidence that decides the matter.

My argument here against the claim that we have knowledge that gives us reason to prefer those hypotheses covering more individual distributions under the Q-division (those holding in more state descriptions),

does not, in fact, materially affect the main point I wish to make about the contradictions: it was acknowledged that the contradictions could be resolved if we used the principle of indifference in relation to relevant evidence. The discussion in the above two paragraphs only concerns what we may legitimately appeal to as relevant evidence. Now the important fact at the moment is that Carnap, whose solution to the contradictions we are considering, does not, and could not, regard the information of the number of state descriptions in which a hypothesis holds as evidence relevant to that hypothesis. Rather the number of state descriptions in which a hypothesis holds determines its range and it is other sentences having different ranges that constitute evidence for the hypothesis. How then can he justify as other than arbitrary the choice of the Q-division as the unique division to which the principle of indifference is to be applied? The probability values that result are different for certain different hypotheses and if we think of the probability values as reflecting degrees of rational credence, such assignments apparently reflect different degrees of rational credence to be accorded various hypotheses. But it is difficult to see how, in the absence of any evidence relevant to two competing hypotheses, we can rationally have more credence in one than the other, and so such probability assignments will be arbitrary.

The arbitrariness of the semantics offered by a consistent assignment of equal probabilities in relation to the null relevant evidence is most clear when we consider geometrical probabilities. We have seen that any inconsistency is avoided in a wholly natural way if we assign equal

probabilities on the basis of equally balanced relevant evidence, such as, for Bertrand's paradox, knowledge of the kind of method used for drawing the chord and knowledge of the proportions yielded by this method. But what are we to do in relation to the null relevant evidence? To ask this is, of course, just to ask which of the three methods of determining probabilities used in Bertrand's paradox is to be chosen as the correct one in a state of total ignorance. The force of Bertrand's paradox is that any choice must be wholly arbitrary - this, I take it, is what Borel and Poincaré meant when, as Keynes puts it, they found the choice of any particular procedure for determining geometric probabilities with the principle simply a matter of 'convention'. Unlike non-geometrical probabilities for which certain restriction to avoid contradictions (such as Carnap's proposal to use only the Q-division) appear at first glance not to be wholly arbitrary, geometrical probabilities are amenable only to restrictions that are all patently arbitrary.

We thus are forced to admit that if we wish the semantics offered by the principle of indifference to include comparative relations between pairs of hypotheses and their null relevant evidence we must accept an arbitrary semantics if our semantics is to provide a consistent and thus adequate interpretation of the axiomatic features of the concept of probability. This consideration seems to me the most compelling of any so far adduced to persuade us to exclude such comparative relations from our semantics and thus abandon the attempt to determine a priori probabilities by use of the principle. This, of course, simply reinforces the point made at the end of the previous

section of this chapter, namely that the principle will be in a form that avoids most serious objections if, but only if, its use is restricted to cases where we have relevant and evenly balanced evidence.

### An Historical Remark

Having seen the numerous difficulties that the principle encounters if, but only if, it is used to assign equal probabilities in the absence of all relevant evidence, it now seems appropriate to note that all along most traditional advocates of the principle, as opposed to its critics, intended it to be used primarily to determine equal probabilities on the unproblematic basis of relevant but equally balanced evidence. Once we have established this it should, I think, be acknowledged that the principle of indifference, even as traditionally advocated, deserves the respect due to the first (if somewhat confused) attempt to define probability as the comparative concept of evidential support, rather than the disrepute that its association with dubious a priori probabilities has brought it.

First of all let us consider Bernoulli's second maxim already discussed, namely that we are to take all the relevant evidence that we can in a given situation. Such a maxim was hardly called for if it were envisioned that the total evidence possible would, in most cases of interest, turn out to be the null relevant evidence. Given the expectation that we shall usually have to deal



with the null relevant evidence, we need not be exhorted to consider the total relevant evidence possible, for there is no such evidence to be considered in part or in toto. Thus Bernoulli's maxim is only really appropriate in the context of the expectation that the total evidence will, on the whole, be relevant.

Admittedly this is not a particularly strong argument for showing that Bernoulli was primarily concerned with comparative relations of evidential support involving relevant evidence; however it does suggest that a careful look at the kind of use that the principle was put to might well reveal a predominant concern among its proponents for such comparative relations. Now proponents of the principle were, somewhat notoriously, given to asserting that the principle could determine probability values on the basis of mere ignorance. But the ignorance spoken of is more often than not further characterized as ignorance of any reason for preference. And, as we have seen, such ignorance of reason for preference can arise from awareness of evidence relevant to the matter at hand, when that evidence is evenly balanced between the alternatives. Judgments of equal probability made in such circumstances are, as we have seen, based on unproblematic comparative relations of evidential support between pairs of evidence and hypothesis where the evidence is relevant to each hypothesis.

Furthermore, let us consider the kinds of probalistic situations to which the principle traditionally was applied. Despite the occasional excursion of proponents of the principle to matters as far from games of chance as theology and astronomy, the principle was most characteristically applied to, and illustrated by, ordinary and simple games of chance involving dice, roulette

wheels, cards etc. Now of course, coins and dice have been tossed for a very long time; the physical properties that are deemed relevant are simple measurable magnitudes of weight and size, while the statistical results, if carefully gathered, can be easily evaluated. (Many philosophical studies of probability actually cite lengthy trials of coin tossing). Moreover, even in the absence of any seriously gathered statistics it is widely believed (and correctly so in my opinion) that the actual proportions occurring in tossing of ordinary coins or dice, as well as spins of roulette wheels and distributions of shuffled cards, are roughly equal for each ordinary alternative, e.g. 'heads' or 'black' etc., involved in these games. And, I think, it is just such a belief that serves as the basis for the judgments of equal probability made by the principle of indifference in these cases. We have no reason to prefer heads to tails in a coin in ordinary circulation because such coins are believed, on the whole, to produce heads in one out of every two trials, similarly so for the other games. And we have seen at length already, the assignment of equal probabilities in such circumstances is clearly based on the existence of relevant but evenly balanced evidence.

With this in mind let us look at the writings of an eminent philosopher who, among English authors, has the dubious distinction of being known as the most ardent advocate of the principle's application on the basis of null relevant evidence, namely W. S. Jevons. Now, admittedly we will find Jevons advocating equal a priori probabilities but, I think, with the above considerations in mind, it will become clear how small a part these a priori probabilities play in his over-all treatment of the principle. This will, I hope,

indicate in general how traditional proponents of the principle primarily regarded it.

Consider the passage from Jevons with which he begins a section called 'Fundamental Principles of the Theory' and which might, at first glance, appear to deal with the assignment of equal probabilities on the basis of the null relevant evidence.

We must treat equals equally, and what we know of one case may be affirmed of every case resembling it in the necessary circumstances. The theory consists in putting similar cases on a par, and distributing equally among them whatever knowledge we possess. Throw a penny into the air and consider what we know with regard to its way of falling. We know that it will certainly fall upon a side, so that either head or tail will be uppermost; but as to whether it will be head or tail, our knowledge is equally divided. Whatever we know concerning head, we also know concerning tail, so that we have no reason for expecting one more than the other. The least predominance of belief to either side would be irrational; it would consist in treating unequally things of which our knowledge is equal.

The theory does not require, as some writers have erroneously supposed, that we should first ascertain by experiment the equal facility of the events we are considering... Nor does the theory show that the coin will fall as often on the one side as the other. It is almost impossible that this should happen, because some inequality in the form of the coin, or some uniform manner in throwing it up, is almost sure to occasion a slight preponderance in one direction. But as we do not previously know in which way a preponderance will exist, we have no reason for expecting head more than tail. Our state of knowledge will be changed should we throw up the coin many times and register the results. Every throw gives us some slight information as to the probable tendency of the coin, and in subsequent calculations we must take this into account. In such cases experience might show that we had been

entirely mistaken; we might expect that a die would fall as often on each of the six sides as on each other side in the long run; trial might show that the die was a loaded one, and falls most often on a particular face. The theory would not have misled us: it treated correctly the information we had, which is all that any theory can do.<sup>11</sup>

Notice that in this passage Jevons repeatedly emphasizes that it is knowledge, as well as ignorance, that is evenly distributed in the application of the principle of indifference – indeed he explicitly states at another point that applications of the principle might as well be characterized as 'engaged with the equal distribution of knowledge' as by saying it is engaged 'with the equal distribution of ignorance'.<sup>12</sup> On the analysis offered in this chapter, we can understand this equal distribution of knowledge in virtue of which equal probabilities are assigned as the awareness of relevant evidence of a kind that is equally balanced between the alternatives. For example, the coin of which Jevons says our knowledge is equally divided between heads and tails is just that kind of case in which, as we have just seen, we do have relevant evidence, namely the information that the coin has been drawn out of circulation in the ordinary fashion and moreover, that the ratio of heads to tails in the past with such coins has been roughly even. That Jevons is concerned with such relevant evidence, evenly balanced by equal proportions in past trials, can be seen in his other example of the die. Here, by the principle of indifference, we are led to expect that 'a die would fall as often on each of the six sides as on each other in the long run.' But surely such an expectation is not formed merely on the basis of null relevant evidence,



past experience has shown that such is the proportion to be found among dice as a whole. That the die, unbeknown to us, might be loaded does not affect the fact that in the absence of information about its exact weighting, we do have relevant evidence about it, namely that it is made out of a certain substance and, to the naked eye, has regular edges with symmetrical markings etc. And, of course, such objects have been seen in the past to produce roughly equal proportions for each face. Further information about its exact weight is, of course, relevant and indeed highly significant but this is not to say that in its absence we lack any relevant information.

Finally let us consider in this light what those interested in the minutiae of the principle's history might well regard as the nadir of a priori probabilities in the pages of Jevon's Principles of Science and the footnotes of Keynes's Treatise. Jevons at one point remarks:

If I ask the reader to assign the odds that a Platythliptic Coefficient is positive he will hardly see his way of doing so, unless he regard them as even.<sup>13</sup>

Even Keynes, who as we have seen was sympathetic to the principle, balks here for he concludes, naturally enough, that this passage shows that Jevons will go so far in assigning equal probabilities in the absence of relevant information as to assign them to even apparently meaningless propositions:

Jevons's particular example, however, is also open to the objection that we do not even know the meaning of the subject of the proposition. Would he maintain that there is any sense in saying that for those who know no

Arabic the probability of every statement expressed in Arabic is even. How far has he been influenced by known characteristics of the predicate 'positive'? Would he have assigned the probability  $\frac{1}{2}$  to the proposition 'A Platythliptic Coefficient is a perfect cube'? What about the proposition 'A Platythliptic Coefficient is allo-geneous'?

But Keynes seems to have missed Jevons's point here. Jevons's example was chosen particularly because the proposition appears meaningless. Moreover, and this is the important point, it was chosen as meaningless in order to show how rare are the cases in which probability values are assigned in the absence of any relevant evidence. Immediately before the passage just quoted from Jevons we find him asserting:

We can hardly indeed invent a proposition concerning the truth of which we are absolutely ignorant, except when we are entirely ignorant of the terms used.

So Jevons's example of the 'Platythleptic Coefficient' far from showing him to be primarily interested in using the principle of indifference in cases where we lack relevant evidence, is produced only to show how rare such cases are (although, of course, there is no doubt that he did think in such cases we should assign the probability  $\frac{1}{2}$  to each proposition and its negation).

Thus save for such cases as the 'Platythliptic Coefficient' and a few literally far fetched cases such as a famous 18th and 19th Century dispute<sup>14</sup> over properties of the star Sirius, even Jevons can be understood as primarily concerned with the assignment of equal probabilities on the basis of equally balanced relevant evidence rather than the null relevant evidence. And so,

as I suggested before, in fairness to many advocates of the principle, we should consider it as primarily intended to offer the plausible definition of probability as the comparative concept of evidential support, where the evidence employed is relevant to the hypotheses in question.

### Subjectivism and The Principle of Indifference

Now I do not wish to claim that all objections to the principle have arisen because of its application in relation to the null relevant evidence. Thus even if we accept the findings of the previous sections that, (1), most objections to the principle do regard its use for determining a priori probabilities and, (2), that the principle was primarily intended by classical writers to determine probabilities on the less problematic basis of relevant but equally balanced evidence, there still may remain certain objections to the principle. In fact the only significant objection to the principle still outstanding is that it leads to a subjective conception of probability and in this section of this chapter I will examine this objection at length on the basis of the interpretation I have put forward of the principle as definitional in character.

Let me first of all explain the sense in which it is not, in my opinion, the use of the principle to determine a priori probabilities that accounts for the appearance of subjectivism often associated with the principle. The important point here is put quite simply in terms of our example of the

Platythliptic coefficient being positive. Here admittedly we have no evidence relevant to this proposition, but this is hardly a subjective matter - it is a quite objective fact that no evidence can be relevant to this proposition for it has no clear meaning. Thus there is nothing 'subjective' about assigning this proposition the probability  $\frac{1}{2}$ , as objectionable as it is on other grounds.

Similarly if it is thought that we have no evidence relevant to a given coin landing heads until it is tossed repeatedly (I, for one, don't think this), any coin which has not yet been tossed will be one for which, quite objectively, there is no relevant evidence.

So whatever grounds there are for claiming that the principle leads to subjectivism, they will not be its admittedly problematic use in relation to the null relevant evidence as such. The real basis for such charges can, I think, be uncovered if we recall the discussion in Chapter V of subjectivism. In Chapter V I indicated that theories which define probability as a relation between evidence and hypothesis could lead to two different kinds of subjectivism: first of all subjectivism could enter into the theory proper, as I put it. A theory of probability as the relation of evidential support would be subjectivist in this sense if the probability value that a person were to assign pairs of evidence and hypothesis on the basis of the theory could depend on his actual beliefs about the relationship of the evidence to the hypothesis. Both Keynes and Carnap were at pains to show that their theories of probability as a logical relation of evidential support made the degree of probability to be assigned to pairs of evidence and hypothesis depend only on purely logical relations between evidence and



hypothesis, which obtained quite independently of what any individual might believe. Thus, as I also said about my own theory of probability as an empirical relation between evidence and hypothesis, no subjectivism entered into these theories proper. But there was a second variety of subjectivism to which theories of probability as the relation of evidential support could lead to and this was one which would come about when these theories were applied to situation with the hope of determining the unique probabilities often needed for the purpose of action. If, as Carnap appeared to propose, the methodological requirement to take the total available evidence was understood in the weak sense that a given person at a given time was only required to consider all the evidence he then knew, the probability values determined by this requirement for purposes of action would depend on the amount of evidence known to each person. Subjectivism of this kind could be avoided for a theory of probability as a relation between evidence and hypothesis if, but only if, the methodological rules by which that theory was applied to yield unique probabilities were formulated in a sufficiently strong manner.

With this analysis of the two different varieties of subjectivism a theory of probability as a relation between evidence and hypothesis can lead to, it is possible to give a perfectly clear account of why the principle of indifference was so often believed to lead to subjectivism. On the analysis I have offered, the principle is an early and rudimentary version of the definition of probability as a relation between evidence and hypothesis - a version in which probability is defined, through fixing identity conditions,

as a comparative relation between evidence and hypothesis, or, as I have often said, defined as the comparative concept of evidential support. As such it can lead to just these two kinds of subjectivism unless proper precautions are taken.

First of all let us consider what I have called subjectivism in a theory proper: here a theory which defines probability as a relation between evidence and hypothesis would lead to subjective results if the probability values assigned to pairs of evidence and hypothesis were made to depend on people's actual beliefs about the relationship between the evidence and hypotheses. In a theory in which probability is defined as a comparative relation of evidential support, pairs of evidence and hypothesis are not assigned numerical probabilities one by one, but only comparative probabilities in relation to other pairs of evidence and hypothesis. In recent times probability theorists have devoted considerably less attention to the comparative concept of probability than to the quantitative concept; however, even since Keynes, Koopman has offered a theory of probability as a comparative relation of confirmation<sup>15</sup> and Carnap himself devotes considerable attention to giving the outlines of such a theory<sup>16</sup>.

Carnap, and this is what is important for us, is again at pains to develop this theory in a way that makes the comparative probabilities assigned to pairs of evidence and hypothesis in relation to each other depend only on purely logical relations between the evidence and hypothesis of each pair. As with quantitative probabilities, comparative probabilities assigned to pairs of evidence and hypothesis will not, in Carnap's system, depend on the subjective matter of individual beliefs about the evidence and hypothesis. Thus at least

Carnap's theory of probability as the comparative concept of confirmation is free from our first variety of subjectivism.

But, of course, there is no necessity to regard the comparative degrees of probability holding between pairs of evidence and hypothesis to depend solely on logical relations between evidence and hypothesis. These comparative probabilities might depend on empirical relations between evidence and hypothesis or, simply, on opinions people have about the evidence and the hypothesis. Certainly in practice comparative degrees of probability are not always assigned on the basis of purely logical relations - people of certain religious persuasions believe hypotheses to be better supported by evidence of Biblical quotations than by the evidence gathered by scientific observations; others, of course, do not. If comparative degrees of probability between different pairs of evidence and hypothesis were in practice always determined by purely logical relations between evidence and hypothesis, such differences of opinion would not occur. However, as this example clearly illustrates, a theory of probability that allows comparative degrees of probability to be assigned on extra-logical considerations runs the risk of sanctioning different comparative probabilities between the same pairs of evidence and hypothesis for different people depending on their different beliefs - and this, of course, is our first variety of subjectivism.

What of the definition of probability as the comparative concept of evidential support which I have claimed is implicit in the principle of indifference? To take the primary case, are the equal probabilities sanctioned

by the principle when the available evidence gives no reason for preference based on purely logical relations between that evidence and the hypotheses in question? Well, the principle of indifference, as I interpret it, only presents in a rudimentary way a semantics for the concept of probability; among other reasons why this definition must be regarded as rudimentary is that the classical theorists did not take the trouble to explain clearly and rigorously the phrase 'reason for preference'. In the absence of an explanation of this expression we can only conjecture what the classical theorists would have regarded as the grounds for saying a body of evidence did or did not give reason for preference. Was it purely logical considerations, say, of the range of the evidence and competing hypotheses that made one hypothesis preferable to another on given evidence? Or could someone rightly hold that a body of evidence made one hypothesis preferable to another just because it looked that way to him? It seems plausible to think that the classical theorists did not hold the latter view, but, alternatively, it would be a gross anachronism to ascribe the former view to them.

It is now possible to understand one reason that the principle appears to some to lead to subjectivism. As we saw, the crucial idiom of 'reason for preference' was to be analyzed in terms of the comparative concept of evidential support - there would be reason to prefer one hypothesis to another if, but only if, the available evidence supported that hypothesis more than the other. It was in light of this analysis of 'reasons for preference' that I concluded that the principle in fact offered the comparative concept of evidential support as a definiens for the concept of probability - thus the failure



of classical writers to elucidate what constituted 'reasons for preference' then is tantamount to a failure to explicate the definiens they offered for the concept of probability. Our brief consideration of Carnap's (outline of a) definition of probability as the comparative concept of confirmation or evidential support, shows that it is possible to give an account of the definiens here in such a way that the comparative degrees of confirmation assigned to different pairs of evidence and hypothesis would depend solely on logical relations between the evidence and hypotheses in question. But as we also saw, it was not necessary to do<sup>so</sup> - failure to make the comparative probabilities depend on purely logical considerations did, however, run the risk of leaving us with a subjective conception of probability. Since classical theorists writing on the principle did not explain what considerations made evidence provide, or fail to provide, reasons for preference (that is, did not explain the comparative concept of evidential support which is the definiens for the concept of probability offered by the principle), it is perfectly possible to believe that they did not hold that purely logical relations between evidence and hypothesis were the relevant ones. It then is distinctly possible that the definiens for the concept of probability offered by the principle is a concept of evidential support such that comparative degrees of evidential support assigned to pairs of evidence and hypothesis are not determined by purely logical relations between evidence and hypothesis. Depending on what other considerations might for them have determined reasons for preference the conception of probability the principle leads to would be subjective. Certainly the conception would be subjective if, along the lines I hypothetically suggested

above, classical theorists believed a person could rightly hold that a body of evidence gave no reason for preferring one hypothesis to another just because it looked that way to him.

Now Carnap has argued that the classical theorists did on the whole believe probability to be an objective, logical, relation between evidence and hypothesis. In this I think he is correct; however, at the moment we are not concerned with their views 'on the whole' but their views on the principle of indifference. In so far as the classical writers were not forthcoming as to what sort of relations between evidence and hypothesis constituted reasons for preference, it remains an open possibility that the principle leads to a subjective conception of probability - the reasons for preference which by the principle determines probabilities might depend on the subjective matter of peoples' beliefs about the evidence and hypotheses in question.

Now my interest here is not so much to charge the principle with subjectivism or vindicate it from such a charge - in fact I think a verdict of not proven is more appropriate - rather I am more concerned with explaining why the principle might appear to some to lead to a subjectivist conception of probability, but not so to others. One reason for this has emerged so far from my analysis of the principle as a rudimentary version of the definition of probability as a relation between evidence and hypothesis - any theory which defines probability as a relation between evidence and hypothesis will lead to subjectivism if it permits probabilities to be assigned to pairs of evidence and

hypothesis in a way that depends on personal opinions rather than objective relations between evidence and hypothesis. Since the classical writers did not explain what sort of relations between evidence and hypothesis were grounds for preferring one hypothesis to another on a given body of evidence - that is, did not explain what relations determined the degree of support or probability of a hypothesis on a given body of evidence - it is an open possibility that the concept of probability defined by the principle leads to just this variety of subjectivism. This is a possibility that has all too eagerly been fastened on to by the principle's critics.

However on the view given here it is not the case that the definition of probability encapsulated in the principle actually leads to subjectivism; rather it is that that definition is given in such a rudimentary way - and expressed in such a vague intuitive idiom - that the kind of elaboration required in a relational theory to avoid such subjectivism just was not provided. It is an open possibility that the relations deemed relevant by classical writers for determining probabilities by the principle were of a kind that do lead to subjectivism; however, it is just as much an open possibility that the relations deemed relevant were of a kind that would avoid subjectivism. Depending on which open possibility one opted for in interpreting it, the principle would or

would not appear to lead to subjectivism. This I take it is just what has happened in the course of the principle's long and controversial history, with proponents and opponents of the principle seizing on different possibilities of those left open. As both possibilities were genuinely left open by lacunae in the classical theorists' treatment of the principle, the appropriate verdict on the charge of subjectivism so far considered seems to me to be 'not proven'.

The variety of subjectivism we have so far considered is that which I called subjectivism in a theory proper. However, as we saw, there is a second variety of subjectivism to which theories of probability as a relation between evidence and hypothesis can lead: that which may arise when the theory is applied in the hopes of obtaining the unique probabilities so often needed for the purposes of action. We have seen how for this purpose most writers who present a theory of probability as a relation between evidence and hypothesis recommend, in one form or another, the maxim that we are to take the total available evidence; although, as we also noticed, different interpretations of what constitutes available evidence exist, ranging from an objective one to a subjective one. If 'available evidence' is understood in the latter subjective sense, the maxim to take the total available evidence leads to results that are only unique for a given person at a given time.—people knowing other amounts of evidence will decide on a different probability value for a hypothesis for the purposes of action. This, as we saw, constituted a second variety of subjectivism and one that could only be prevented by strong enough methodological rules of application.



Now it was explained near the end of the previous chapter how the principle of indifference would frequently yield unique numerical probabilities. Such results were arrived at because, although encapsulating a comparative and relational definition of probability, the principle was so formulated that any given person at any given time could only apply it in relation to one body of evidence. Just what body of evidence this was turned out to be a matter of some dispute - Mackie and Kneale held that the principle was properly applied by someone on the basis of the evidence he then was aware of, which, of course, meant the total evidence he was then aware of. Blackburn and Bernoulli held that the evidence to be used was the total amount one should or could obtain in a given situation. On either line of interpretation one person at one time could apply the principle on the basis of one body of evidence alone and so would arrive at just one ordering of the competing hypotheses in respect of comparative degrees of probability. This in turn lead to a single numerical value in those cases in which a set of mutually exclusive and exhaustive hypotheses were found equally likely.

It should be clear that if the Mackie-Kneale interpretation of the principle is accepted, the principle can lead to different numerical values of probability for different people in the same situation, depending on the different amounts of evidence they know. Thus, on their interpretation of it, the principle leads to a kind of subjectivism by warranting different probability values for different people in the same situation - different people will apply the principle in relation to different bodies of evidence. Significantly enough, the kind of

subjectivism here is just the second variety we have discussed - the subjectivism that results when a theory of probability as the relation of evidential support is applied to determine unique probabilities on the basis of the evidence which happens to be known at the time by those who are applying the theory.

Conversely, if the Blackburn-Bernow<sup>III</sup> interpretation of the principle is accepted, the principle will not lead to this kind of subjectivism, for on their interpretation the principle is only properly applied in relation to the totality of evidence one could or should have known in a given situation. Both the evidence we could have known and the evidence one should have known are totalities of evidence fixed quite independently of what evidence is actually known by particular people in a situation. Thus if we determine probabilities by applying the principle on the basis of either such totalities of evidence, the results of applying it will not vary because different people happen to know different amounts of evidence. And it was just the possibility <sup>of</sup> such variation in the results of applying the principle to determine unique probabilities that made the principle appear to lead to our second variety of subjectivism.

Now as before I am not primarily interested in arguing for the Kneale-Mackie interpretation of the principle, and so convicting it a form of subjectivism, or arguing for the Blackburn-Bernow<sup>III</sup> interpretation and vindicating it. Rather, again, I am interested in elucidating the reason why the principle has seemed for some to lead to a subjective conception of probability and not seemed so to others. A second reason has now emerged for this - although encapsulating a relational definition of probability, the principle will yield the unique probabilities we so often desire for it is

expressed in such a way that it will always be applied by a person in relation to just one body of evidence. Depending on how the principle is interpreted, this body will either be the totality known to the person who wishes to apply it or a larger totality incorporating evidence which in some sense is available, although not necessarily known to the person in question. On the former interpretation found in Kneale and Mackie, the principle will lead to that kind of subjectivism we diagnosed as arising when we attempt to give unique application to theories which define probability as a relation of evidential support. On the latter interpretation found in Bernoulli and Blackburn this variety of subjectivism is entirely avoided.

Now it might seem in all fairness to Bernoulli we should regard his interpretation of the principle of indifference as correct and thus vindicate the principle here of the charge of subjectivism. But what seems to me the important point is that fair or unfair, the Kneale-Mackie interpretation of the principle has as a matter of fact often been given - and not merely by its critics - and so not surprisingly the principle has appeared to many to lead to subjectivism. That it has not so appeared to others is, of course, explicable given the existence of the alternative Bernoulli-Blackburn interpretation of it.

Thus we find that the principle of indifference as a rudimentary version of a theory in which probability is defined as a relation between evidence and hypothesis can lead to both of what I called subjectivism in a theory proper and subjectivism in the application of that theory. However, it only does so on certain possible, but far from necessary, interpretations of it. In the case

of the former variety of subjectivism, the relevant different interpretations are over what classical writers would have regarded as the grounds for a given body of evidence providing reason for preference and the possibility of different interpretations here stem from *lacunae* in their development of the principle. With the latter kind of subjectivism the different interpretations pertain to exactly what amount of evidence the principle is correctly applied in relation to and no matter what the origin of the different interpretations of the principle over this point, there can be no doubt that such different interpretations do exist and so would lead to conflicting opinions about the principle.

Let me conclude this section by comparing my rather general comments on subjectivism and the principle of indifference with one specific criticism of the principle as leading to subjectivism, that voiced by Kneale. As we have seen Kneale claimed that the principle of indifference led to subjectivism because it allowed probabilities to be assigned in the 'absence of knowledge'. Kneale elaborated this point by saying that the principle of indifference allows us to 'call alternatives equiprobable if we do not know that the available evidence provides a reason for preferring any one to any other'. Blackburn finds this formulation obscure and remarks apropos of Kneale that

...it is not clear why the state of ignorance which we are trying to characterise...is best described in terms of ignorance of the properties of the available evidence. What is this concept - itself very unclear - doing in a description of a state of ignorance about two alternatives?<sup>17</sup>



Although I think Blackburn is correct in remarking on the obscurity of this passage in Kneale, I think it is possible to explain even this objection to the principle within the framework I have provided. How, we may ask, would it be possible for us not to know of the available evidence that it gives reason for preference? It seems to me this is only possible in three ways: first of all we might not actually know all of the available evidence, for then, of course, we could not know if that evidence did give reason for preference. Secondly, assuming we were aware of all the available evidence, we would not know that it gave reason for preference if we had mistakenly evaluated part or all of it. And thirdly, even if we were fully aware of all the available evidence and made no definite errors of misjudgment, we might simply be unable to evaluate some, or all, of that evidence.

. In the first case - failure to be aware some material part of the total available evidence - we clearly would be guilty of determining probabilities subjectively if we accorded equal probabilities to each alternative for we would have failed to consider all the available evidence. This is exactly the second variety of subjectivism I have discussed above and thus, so far, Kneale's charge of subjectivism is just an obscurely stated version of a charge I considered at length. To put the matter generally, it now appears that 'the absence of knowledge' that Kneale (and others) found an objectionable basis for determining probabilities might simply consist in the failure of the person assigning the probabilities to consider the total available evidence and this is a possibility which has been carefully examined already.

A similar point can be made in regard to the next way in which we could, as Kneale puts it, fail to know that the available evidence gives reason for preference - those cases in which we have considered all the available evidence but misjudged it seriously. Obviously probability judgments made on such a basis would be subjective, due to our own failure in valuation. Here probabilities are again objectionably assigned in the absence of knowledge but, this absence of knowledge consists in the absence of someone's knowledge of the genuine import of the available evidence. In so far as the principle of indifference warrants the assignment of probabilities in the absence of knowledge of the genuine import of a body of evidence, it is guilty of the first variety of subjectivism I discussed. Theories of probability could lead to this variety of subjectivism, as the reader will recall, by allowing probability values to be assigned to pairs of evidence and hypothesis in a way that depended on particular peoples' beliefs about the relationship between the evidence and hypothesis in question. If after considering the total available evidence and deciding that in our opinion it did not give reason to prefer any of a set of competing hypotheses - when, objectively speaking, it did give reason for preference - we assign equal probabilities by use of the principle, these equal probabilities would not be based on objective (for example purely logical) relations between the evidence and the hypotheses, but rather would be based on our (subjective) opinions about the relationship of the evidence to the hypothesis. It was exactly such a possibility that as I said was left open by the classical writers' failure to explain clearly their concept of 'reason for preference' and, again, the variety of subjectivism this could lead the principle into has already been dealt with thoroughly in this section.

The third way in which we might not know that the total available evidence gave reason for preference is if we were unable to evaluate some or all of the total available evidence, even though we were aware of it all. The subjectivism that would arise if we allowed the assignment of equal probabilities on the basis of such ignorance can, in fact, be explained in just the way we have explained the subjectivism that would arise if we allowed the assignment of equal probabilities on the basis of misjudging the total available evidence. If we can not evaluate some or all of the total available evidence, the sensible thing to do, of course, is not to make any assignment of probability. If, instead, in this state of ignorance we use the principle of indifference to assign equal probabilities to each of our competing hypotheses, we may just be lucky enough to have hit on the correct probability value for the hypotheses in relation to the total available evidence properly evaluated. However, this would only be a matter of luck and quite often the beliefs about the probability values we would arrive at in this way would not correspond to what would have been arrived at had the total available evidence been properly evaluated. In such cases the probability values we believe in for our hypotheses in light of the total available evidence again obviously do not correspond to any objective relations between the evidence and our hypotheses. The equal probabilities we believe in will simply reflect our own personal beliefs arising from an inability to evaluate all or some of the evidence in question. Moreover the probability values people would come to believe in by using the principle in this manner would vary from individual to individual, depending on their ability to evaluate the available evidence. I take it something very

much like this is one of the possibilities Kneale had in mind when he claimed that the principle led to subjectivism by allowing us to assign equal probabilities when we do not know that the available evidence gave reason for preference. But the subjectivism that the principle leads to here is just our first variety of subjectivism all over again, for with this 'subjectivism in a theory proper' the probability values for pairs of evidence and hypothesis warranted by a theory reflect individual beliefs about these values, rather than objective relations between evidence and hypothesis and this is just the difficulty here.

Thus each kind of subjectivism which would arise from assigning equal probabilities when 'we do not know that the available evidence gives reason for preference' is an instance of one or the other of the two varieties of subjectivism considered earlier. Thus Kneale's somewhat obscure formulation of the charge of subjectivism against the principle in fact readily accords with the explanation given in this section of why, on certain interpretations of it, the principle appeared to lead to subjectivism. But, let me finally add, we also have seen that there are possible, and indeed plausible, interpretations of the principle that vindicate it from the charge of subjectivism of either variety and so, of course, vindicate it from Kneale's obscure formulation of such charges.



### Summary of Discussion of The Principle

I hope that this discussion of the charges of subjectivism levelled against the principle will confirm in the readers mind the over-all interpretation to the principle espoused here. The principle is seen as expressing implicitly and in a rudimentary way the definition of probability as the comparative concept of evidential support - theories in which probability is so defined as a relation between evidence and hypothesis can lead to subjectivism in two different ways and thus, as just explained, the principle can well appear to lead to subjectivism in one or both of two ways. However, the principle need not be understood in a way that leads to either variety of subjectivism as there exist other plausible interpretations of the principle that do not lead to any kind of subjectivism. Thus with the analysis of the principle as largely definitional in character I hope to have shown both why charges of subjectivism have been raised against it and also how the principle can be interpreted so as not to lead to subjectivism. Similarly, we saw previously that if the principle is regarded as providing, in a rudimentary way, semantics for the concept of probability in terms of comparative relations of evidential support, other controversies surrounding the principle can be perspicuously understood and to a large extent resolved: in the semantics offered by the principle equal degrees of probability are equated with equal degrees of evidential support. It then turned out that evidence could provide equal degrees of evidential support for competing hypotheses in two quite different ways: the evidence could be relevant to the hypotheses, yet equally balanced, or the evidence could be totally irrelevant.

That is to say, the semantics offered by the principle involved two distinct kinds of evidentiary relations, that between hypotheses and evidence relevant to them and that between hypotheses and evidence irrelevant to them. The plausibility of the semantics offered by the principle in terms of evidentiary relations of the former kind may well tempt us to accept a semantics including the latter kind as well, but, as we saw, this involves numerous difficulties not the least of which was the contradictions the principle has been found to lead to. Such considerations suggested that the principle would be far more satisfactorily formulated if restricted to prevent the assignment of a priori probabilities; moreover it was found on careful analysis that even an ardent defender of the principle's capacity to assign a priori probabilities such as Jevons was, in fact, primarily concerned with equal probabilities assigned on the basis of relevant but equally balanced evidence. As this was the primary concern of even the classical theorists, it seemed proper to regard the principle as primarily intended to offer the highly plausible semantic definition of probability as the comparative concept of evidential support, where the evidence which supports the hypotheses is understood as relevant to it.

Since the definition of probability explicitly given by classical theorists - the ratio of favourable to (equi) possible cases - is to be regarded as an axiomatic definition of probability for which the principle of indifference provided part of a semantic interpretation, the classical theory as a whole can, in this way, be regarded as a first and therefore not surprisingly rudimentary attempt to state what we would now regard as a highly plausible theory of

probability, with an axiomatic and semantic part. There can be no doubt, of course, that the semantic part expressed in the principle is quite rudimentary, and in places egregiously incomplete - vide, the failure of the classical theorists to elucidate the concept of 'reason for preference'. Moreover the semantic part expressed in the principle is admittedly somewhat confused being ambiguous between warranty assignments of equal probability based on relevant and equally balanced evidence versus their assessment on the null relevant evidence. Such lacunae and ambiguities are what have lead to the apparently interminable controversies over the principle but such difficulties should not prevent us from seeing the principle for what it is: the first attempt to put forward the highly plausible definition of probability as the relation of evidential support. It is as such that the principle has, I believe, rightly commanded a central place in the history of probability theory.

As it is perhaps fitting to conclude the last chapter of this thesis with a few general remarks, we may also note that this concept of probability as a relation between evidence and hypothesis first put forward by the principle of indifference has, elsewhere in this thesis, been found in varying forms to underlie a variety of apparently quite different theories of probability. In Chapter I - III I argued that the proper account of the idea of statistical probability was that given in my frequency theory for the single case. This was an admittedly relational theory of probability, but in Chapter IV, we considered at length a methodological rule which could give this theory unique application when desired. On the analysis given in Chapter V, this frequency

theory for the single case was seen to be one in which probability was defined as a quantitative and empirical relation between evidence and hypothesis, a relation the quantity of which, so to speak, was fixed by the empirical facts of long run frequencies. This, as we also noted in Chapter V, allowed us to reduce Carnap's apparently distinct concepts of probability<sub>1</sub> and probability<sub>2</sub> to a more fundamental one of probability as a (quantitative) relation between evidence and hypothesis, the quantity of which could be determined by either logical or empirical considerations. The discussion of these last two chapters has shown the classical theory with the principle of indifference to have been concerned with a closely related concept, that of probability as a comparative relation between evidence and hypothesis.



## Notes for Chapter VII

1. Reichenbach, H., op. cit., p. 358.
2. Mellor, H.D., op. cit., p. 138.
3. Keynes, J.M., op. cit., p. 86.
4. Kneale, W., op. cit., p. 186.
5. Keynes, J.M., op. cit., Chap. IV. Unless otherwise stated all subsequent references to Keynes in this Chapter are, again, drawn from his Chapter IV.
6. Mellor, H.D., op. cit., p. 142.
7. Ayer, A.J., op. cit., p. 35.
8. Watling, J.L., 'Confirmation Discomforted' *Revue Internationale de Philosophie*, 1963.
9. Sigwart, C., *Logic*, English edition 1895, vol. ii, p. 220.
10. Blackburn, S.W., op. cit., p. 122.
11. Jevons, W.S., The Principles of Science, Dover edition, New York, 1958, pp. 201-202.
12. Ibid., p. 199.
13. Ibid., p. 213.
14. Ibid., and Keynes, J.M., op. cit., p. 172.
15. Koopman, B.O., *The Axioms and Algebra of Intuitive Probability*, *Ann. Math. Ser. 2*, 41 (1940) pp. 269-292.
16. Carnap, LFP, Chap. VII.
17. Blackburn, S.W., op. cit., pp. 119-120.

## BIBLIOGRAPHY

- AYER, A. J., Probability and Evidence, Macmillan, London 1972.
- AYERS, M., The Refutation of Determinism, Methuen, London 1968.
- BERNOULLI, J., Ars Conjectandi, Basil, 1713.
- BLACKBURN, S. W., Reason and Prediction, Cambridge Univ. Press, Cambridge 1973.
- CARNAP, R., The Continuum of Inductive Methods, University of Chicago Press, Chicago 1952.
- The Logical Foundations of Probability, 2nd ed., University of Chicago Press, Chicago 1962.
- Testability and Meaning, Reprinted in Readings in the Philosophy of Science, ed. Feigl and Brodbeck, Appleton-Century-Crofts, New York, 1953, pp. 47-88.
- CHAITIN, G. J., Randomness and Mathematical Proof, Sci. Am. vol. 232, no. 5, 1975, pp. 47-52.
- CHURCH, A., On the Concept of a Random Sequence, Bull. Am. Math. Soc. vol. 46, 1940, pp. 130-135.
- FINE, T., Theories of Probability, Academic Press, New York, 1974.
- FREGE, G., The Foundations of Arithmetic, Trans. J. L. Austin, 2nd rev. ed., Basil Blackwell, Oxford 1968.

- GILLIES, D.A., An Objective Theory of Probability, Methuen, London 1973.
- HACKING, I., The Emergence of Probability, Cambridge Univ. Press, Cambridge 1975.
- The Logic of Statistical Inference, Cambridge Univ. Press, Cambridge, 1965.
- HEMPEL, C.G., Maximal Specificity and Lawlikeness in Probabilistic Explanation, *Phil Sci* vol. 35, 1968, pp. 116-33.
- JEFFREYS, H., Theory of Probability, 3'd ed. Oxford Univ. Press, Oxford 1961.
- JEVONS, W.S., The Principles of Science, Dover ed. New York, 1958.
- KEYNES, J.M., A Treatise on Probability, Macmillan, London 1921.
- KNEALE, W., Probability and Induction, Oxford Univ. Press, Oxford, 1949.
- KYBURG, H.E., More on Maximal Specificity, *Phil. of Sci.* vol. 37, 1970, pp. 29-300.
- Probability and Randomness, *Theoria*, vol. 29, 1963, pp. 27-55.
- LOVELUND, D.W., A New Interpretation of the Von Mises Concept of a Random Sequence, *Zeit. Math Logik und Grundlagen der Math.*, vol. 12, 1966, pp. 277-294, reviewed *JSL*, vol. 38, p. 357.
- LUCAS, J.R., The Concept of Probability, Clarendon Press, Oxford 1970.

- MACKIE, J. L., Truth Probability and Paradox, Clarendon Press, Oxford 1973.
- MARTIN-LOF, P., The Definition of Random Sequences, *Information and Control*, vol. 9, 1966, pp. 602-19.
- MELLOR, D. H., The Matter of Chance, Cambridge Univ. Press, Cambridge 1971.
- Von MISES, R., Probability, Statistics and Truth, 2nd rev. ed., George Allen Univers., New York, 1957.
- POPPER, K., The Logic of Scientific Discovery, Revised ed., Hutchinson, London 1968.
- REICHENBACH, H., The Theory of Probability, University of California Press, Berkeley, 1949.
- SALMON, W. C., The Foundations of Scientific Inference, University of Pittsburgh Press, Pittsburgh, 1967.
- SIGWART, C., Logic, English ed., Freiburg 1895.
- VENN, J., The Logic of Chance, London 1866.
- WAISMANN, F., Logische Analyse des Wahrscheinlichkeitsbegriffs, *Erkenntnis* I, 1930-31, pp. 228-248.
- WATLING, J., Chance, *Arist. Soc. Suppl.* vol. 73, 1969, pp. 37-48.
- Confirmation Discomforted, *Revue Internationale de Philosophie* 1963, vol. 64, pp. 155-170.
- WITTGENSTEIN, L., Tractatus Logico - Philosophicus, transl. Pears, D. F. and McGuinness, B. F., Routledge & Kegan, London, 1961.



